

NORBERTO RITZMANN JÚNIOR

**Método para Otimização de Janelas de Tempo e
Discretização para Classificação de Movimentos
Futuros de Ações da Bolsa de Valores**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Curitiba

2016

NORBERTO RITZMANN JÚNIOR

**Método de Otimização de Janelas de Tempo e
Discretização para Classificação de Movimentos
Futuros de Ações da Bolsa de Valores**

Dissertação de Mestrado apresentada ao Programa de
PósGraduação em Informática da Pontifícia Universidade
Católica do Paraná como requisito parcial para obtenção do
título de Mestre em Informática.

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Júlio Cesar Nievola

Curitiba

2016

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central

Ritzmann Júnior, Norberto
R615m Método para otimização de janelas de tempo e discretização para
2016 classificação de movimentos futuros de ações de bolsas de valores / Norberto
Ritzmann Júnior; orientador: Júlio Cesar Nievola. – 2016.
82 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná,
Curitiba, 2016
Bibliografia: f. 69-71

1. Mineração de dados (Computação). 2. Bolsa de valores. 3. Algoritmos
genéticos. 4. Mercado futuro – Processamento de dados. I. Nievola, Júlio
Cesar. II. Pontifícia Universidade Católica do Paraná. Programa de
Pós-Graduação em Informática. III. Título.

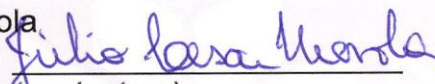
CDD 22. ed. – 006.312

**ATA DE DEFESA DE DISSERTAÇÃO DE MESTRADO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**

DEFESA DE DISSERTAÇÃO DE MESTRADO Nº 19/2016

Aos 23 dias do mês de Março de 2016 realizou-se a sessão pública de Defesa da Dissertação "**Método de Otimização de Janelas de Tempo e Discretização Para Classificação de Movimentos Futuros de Ações da Bolsa de Valores**" apresentado pelo aluno **Norberto Ritzmann Júnior**, como requisito parcial para a obtenção do título de Mestre em Informática, perante uma Banca Examinadora composta pelos seguintes membros:

Prof. Dr. Júlio Cesar Nievo
PUCPR (Orientador)


(assinatura)

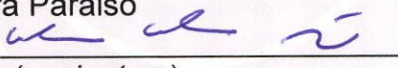
APROVADO
(Aprov/Reprov)

Prof. Dr. Alceu Britto Junior
PUCPR


(assinatura)

APROVADO
(Aprov/Reprov)

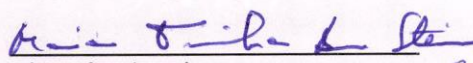
Prof. Dr. Emerson Cabrera Paraiso
PUCPR


(assinatura)

APROVADO
(Aprov/Reprov)

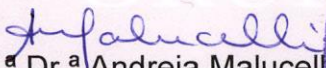
Prof.^a Dr.^a Maria Teresinha Arns Steiner

PUCPR

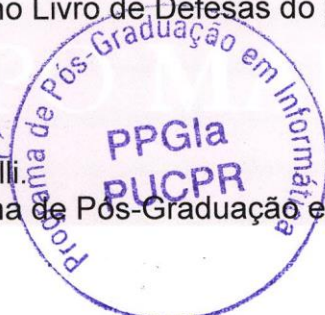

(assinatura)

APROVADO
(Aprov/Reprov)

Conforme as normas regimentais do PPGIa e da PUCPR, o trabalho apresentado foi considerado APROVADO (aprovado/reprovado), segundo avaliação da maioria dos membros desta Banca Examinadora. Este resultado está condicionado ao cumprimento integral das solicitações da Banca Examinadora registradas no Livro de Defesas do programa.


Prof.ª Dr.ª Andreia Malucelli

Coordenadora do Programa de Pós-Graduação em Informática.



*Dedico este trabalho à minha esposa que muito
dedicou para que tudo fosse possível*

Sumário

Lista de Figuras.....	ix
Lista de Tabelas.....	xi
Lista de Abreviaturas	xiii
Resumo.....	xv
Abstract.....	xvi

Capítulo 1

Introdução.....	1
1.1 Motivação.....	2
1.2 Problema	3
1.3 Objetivos	4
1.4 Hipótese.....	4
1.5 Organização do Documento	5

Capítulo 2

Fundamentação Teórica.....	6
2.1 Mercado de Ações	6
2.2 Previsão em Séries Temporais	7
2.3 Características da Previsão em Séries Temporais Financeiras	8
2.3.1 Atributos dos Preços das Ações	9
2.3.2 Indicadores Financeiros	10
2.3.2.1 Indicadores Baseados em Tendência	10
2.3.2.2 Indicadores Momentum.....	11
2.3.2.3 Indicadores Baseados em Volume	11
2.3.2.4 Indicadores Osciladores	11

2.3.2.5 Moving Average (MA)	11
2.3.2.6 Commodity Channel Index (CCI)	13
2.3.2.7 Relative Strength Index (RSI)	13
2.3.2.8 Oscilador Estocástico	14
2.3.2.9 Moving Average Convergence/Divergence (MACD)	15
2.3.2.10 Accumulation/Distribution (A/D)	16
2.3.2.11 Williams %R	16
2.3.2.12 Psychological Line (PSY)	17
2.3.2.13 Indicadores Direcionais	17
2.3.2.14 Indicadores Aroon	19
2.3.2.15 Transformação de Indicadores em Viés e Osciladores	20
2.4 Descoberta do Conhecimento em Séries Temporais Financeiras	21
2.4.1 Métodos de Classificação	22
2.4.2 Técnicas de Redução de Dimensionalidade	24
2.4.2.1 Seleção de Atributos	24
2.4.3 Técnicas de Discretização	27
2.4.4 Algoritmos Genéticos	28
2.5 Considerações	30
 Capítulo 3	
Estado da Arte	31
3.1 Considerações	37
 Capítulo 4	
Método	38
4.1 Características dos Dados	38
4.2 Algoritmo para Otimização de Janelas de Tempo	42
4.3 Algoritmo para Otimização de Atributos Discretos	48
4.4 Configuração dos AGs	50
4.5 Avaliação dos Resultados	51

4.5.1 <i>Taxa de Acerto</i>	51
4.5.2 <i>Algoritmo de Negociação</i>	51
4.6 Base de Dados	53
4.7 Considerações	53
 Capítulo 5	
Resultados Obtidos	54
5.1 <i>Algoritmo Genético para Otimização de Janelas de Tempo</i>	54
5.2 <i>Resultados do Processo Completo</i>	57
 Capítulo 6	
Conclusão e Trabalhos Futuros	64
6.1 Conclusão	64
6.2 Trabalhos Futuros	66
 Referências Bibliográficas	67
Apêndice	70

Lista de Figuras

Figura 1.1: Exemplo de tendência de baixa com muitos ruídos	3
Figura 2.1: MACD e Oscilador MACD	15
Figura 2.2: Indicadores de Direcionais, ação Petrobrás (PETR4) de 09/14 - 02/15	19
Figura 2.3: Indicadores Aroon, ação Petrobrás (PETR4) de 09/14 - 02/15	20
Figura 2.4: Pseudo código do algoritmo ReliefF (Kononenko 1994)	27
Figura 2.5: Algoritmo genético básico	29
Figura 4.1: Histograma dos atributos Williams %R, Estocástico %K e ADX	39
Figura 4.2: Cruzamento dos atributos ADX com Estocástico %K e ADX com Williams %R	39
Figura 4.3: Superfícies de decisão dos algoritmos CART <i>Decision Tree</i> , <i>Random Forest</i> , Gaussian <i>Naïve</i> -Bayes	40
Figura 4.4: Superfícies de decisão dos algoritmos <i>k-Nearest Neighbors</i> , <i>Linear</i> SVM, RBF SVM e <i>Polynomial</i> SVM	41
Figura 4.5: Exemplo de utilização da variável n na fórmula SMA	42
Figura 4.6: Primeira etapa - AG para otimização de janelas de tempo	44
Figura 4.7: Gráfico de resultados rodados com <i>Random Forest</i> 80 árvores variando apenas o parâmetro n do Algoritmo ReliefF	47
Figura 4.8: Segunda etapa, otimização de discretização	49
Figura 4.9: Etapas 1 e 2 integradas	50
Figura 4.10 - Probabilidade de distribuição em árvores de decisão	52
Figura 4.11: Divisão da base de dados	53

Figura 5.1: Histórico de negociações utilizando AGOJT para o índice S&P 500 comparado com C&M	61
Figura 5.2: Histórico de negociações utilizando o processo completo para o índice S&P 500 comparado com C&M.....	61
Figura 5.3: Histórico de negociações utilizando AGOJT para a ação Microsoft comparando com C&M	61
Figura 5.4: Histórico de negociações utilizando o processo completo para a ação Microsoft comparado com C&M.....	61
Figura 5.5: Histórico de negociações utilizando AGOJT para a ação Exxon Mobil comparado com C&M.....	62
Figura 5.6: Histórico de negociações utilizando o processo completo para a ação Exxon Mobil comparado com C&M	62

Lista de Tabelas

Tabela 3.1: Indicadores técnicos selecionados (Chang, Fan, e Lin 2011).....	33
Tabela 3.2: Indicadores técnicos selecionados para cada ação (Chang, Fan, e Lin 2011)	34
Tabela 3.3: Resultados do método LSSVM comparados à PNN, QDA e LDA (Wang e Shang, 2014).....	36
Tabela 3.4: McNamar (<i>p-values</i>) para comparação de performance	36
Tabela 3.5: Comparativo SVM, VW-SVM, SVM + F-score e C&M	37
Tabela 4.1: Indicadores técnicos que terão a janela de tempo otimizados	46
Tabela 5.1: Atributos selecionados e tamanhos de janelas para Microsoft.....	55
Tabela 5.2: Atributos selecionados e tamanhos de janelas para Exxon Mobil	55
Tabela 5.3: Atributos selecionados e tamanhos de janelas para S&P 500.....	55
Tabela 5.4: Resultados AGOJT e comparações.....	56
Tabela 5.5: Resultados do algoritmo de negociação de AGOJT comparado C&M	56
Tabela 5.6: Tamanhos de janela padrão comparados com C&M.....	57
Tabela 5.7: Precisão no processo completo, AGOJT e AGOAD	58
Tabela 5.8: Resultados de negociação no processo completo, AGOJT e AGOAD.....	58
Tabela 5.9: Formato dos atributos encontrados no AGOAD para S&P 500	58
Tabela 5.10: Formato dos atributos encontrados no AGOAD para Microsoft.....	59
Tabela 5.11: Formato dos atributos encontrados no AGOAD para Exxon Mobil	59
Tabela 5.12: Comparação dos resultados.....	60
Tabela 5.13: Sumarização das operações realizadas entre lucros e prejuízos	63
Tabela Apêndice 1: Negociações executadas pelo algoritmo final na ação Microsoft	70

Tabela Apêndice 2: Negociações executadas pelo algoritmo final na ação Exxon Mobil .. 73

Tabela Apêndice 3: Negociações executadas pelo algoritmo final no índice S&P 500 77

Lista de Abreviaturas

A/D	<i>Acummulation/Distribution</i>
AG	Algoritmo Genético
AGOJT	Algoritmo Genético para Otimização de Janelas de Tempo
GOAD	Algoritmo Genético para Otimização de Atributos Discretos
AM	Aprendizagem de máquina
ADX	<i>Average Directional Index</i>
C&M	Compra e Manutenção
CCI	<i>Commodity Channel Index</i>
CFS	<i>Correlation-based Feature Selection</i>
EMA	<i>Exponential Moving Avarage</i>
EFD	<i>Equal Frequency Discretization</i>
EWD	<i>Equal Width Discretization</i>
HME	Hipótese do Mercado Eficiente
IA	Inteligência Artificial
KDD	<i>Knowledge Discovery in Databases</i>

MA	<i>Moving Average</i>
MACD	<i>Moving Average Convergence/Divergence</i>
-DI	<i>Minus Directional Indicator</i>
-DM	<i>Minus Directional Movement</i>
+DI	<i>Plus Directional Indicator</i>
+DM	<i>Plus Directional Movement</i>
FDT	<i>Fuzzy Decision Trees</i>
GAFD	<i>Genetic Algorithm Feature Discretization</i>
OSCP	<i>Price Oscillator</i>
PSO	<i>Particle Swarm Optimzation</i>
PSY	<i>Psychological Line</i>
Relief	<i>Recursive Elimination of Features</i>
RNA	Rede Neural Artificial
RSI	<i>Relative Strength Index</i>
SMA	<i>Simple Moving Average</i>
SVM	<i>Support Vector Machine</i>
TR	<i>True Range</i>
WMA	<i>Weighted Moving Average</i>

Resumo

A previsão de movimentos futuros das ações é considerada uma tarefa desafiadora devido à sua natureza complexa. Para competir em um mercado que cada vez mais vem utilizando ferramentas para negociação automática de ações, é importante para os investidores institucionais ou pessoas físicas utilizarem ferramentas competitivas. Este trabalho procura através do estudo de técnicas de mineração de dados alcançar um modelo de classificação diária para ações da bolsa Dow Jones e para o índice S&P 500, classificando o movimento do dia seguinte em "alta" ou "baixa". O modelo de classificação construído é combinado com um algoritmo genético para otimização de janelas de tempo dos indicadores técnicos utilizados como atributos iniciais e de um algoritmo genético para otimizar a discretização de atributos. O método proposto utiliza como algoritmo de classificação o algoritmo *Random Forest* e o método de seleção de atributos ReliefF. Os resultados apontam para uma precisão de acerto de 56,83% média e 70,57% de lucro acima da média de mercado.

Palavras-Chave: *Random Forest*, Algoritmos Genéticos, Mineração de Dados, Bolsa de Valores.

Abstract

The future movements forecasting of stocks is considered a challenging task due to its complex nature. For competing in a market that increasingly has been using tools for automated stock trading, it is important to institutional investors or individuals to use competitive tools. This study aimed through the study of data mining techniques to achieve a daily classification model for shares of stock Dow Jones and the S&P 500 index, classifying next day the movement in "up" or "down". The built classification model combines a genetic algorithm for optimization of time windows length of technical indicators used as initial features in the process and a genetic algorithm to optimize the feature discretization. The proposed method used Random Forest as classification algorithm and ReliefF for feature selection. The results shows 56.83% of precision in average and 70.57% of profit in negotiation algorithm over the market average.

Keywords: Random Forest, Genetic Algorithms, Data Mining, Stock Market.

Capítulo 1

Introdução

A previsão de movimentos nos mercados de ações baseado em suas informações históricas era pouco creditada até a década de 1980, devido à Hipótese do Mercado Eficiente – HME (Fama 1970). Na época, Fama ganhou notoriedade ao avaliar que os mercados são eficientes e já precificam notícias, eventos corporativos e macroeconômicos; dificultando que outros possam ter vantagem com o fato. Outras teorias apontaram para esta visão, como a teoria do *Random Walk*, que diz que o preço das ações é influenciado por dois fatores: a ação imediata dos compradores e vendedores, e pelas notícias. Contudo, outros estudos empiricamente testados mostraram que a HME não reflete inteiramente o mercado.

Para a tomada de decisões no mercado financeiro, os operadores de mercado usam principalmente dois tipos de análises. A primeira é a análise fundamental e a segunda é a técnica. A análise fundamental é baseada no estado geral da economia, da indústria e na situação financeira da empresa. Em contrapartida, a análise técnica é baseada nos dados históricos e nos gráficos, partindo do princípio de que o histórico se repetirá. Ou seja, se o histórico do passado for analisado, as tendências futuras poderão ser previstas. A análise técnica ainda se subdivide em outros dois métodos: indicadores baseados em estatística e padrões gráficos. Ambos os métodos são utilizados em conjunto, sendo que o primeiro método utiliza indicadores estatísticos, como as médias móveis, para encontrar relações matemáticas no histórico a ser utilizado na previsão de valores futuros no mercado. E, através do segundo método, é verificado se um padrão específico aparece em um gráfico para então prever as tendências futuras (Bagheri, Mohammadi Peyhani, e Akbari 2014).

A aplicação de mineração de dados neste contexto pode conter o objetivo de prever um valor futuro de uma ação, normalmente utilizando regressão de valores. A mineração de dados também é aplicada para descobrir uma tendência de valor futuro, descobrindo assim os momentos de entrada, saída e manutenção de uma ação. Além disso, encontram-se pesquisas no sentido de construir uma carteira de ações que maximizam o lucro e minimizam o risco. Outros objetivos comuns também são: extração de regras, agrupamento por características comuns a fim de encontrar uma ação líder que indique alta para o grupo inteiro de ações e elaboração de algoritmos que sinalizam para o momento de compra ou venda de uma ação.

1.1 Motivação

As quedas nas taxas de juros internacionais, que ocorreram após a crise de 2008, principalmente em países desenvolvidos, onde estas taxas chegaram a ficar abaixo de 1%, tornaram o investimento em ações mais atrativos que anteriormente. Com isso, fundos de investimento e fundos de pensão passaram a investir mais em ações, ampliando o interesse na utilização de técnicas e algoritmos para administrar o risco e maximizar a possibilidade de ganhos. Contudo, nas pesquisas mais recentes que utilizam indicadores técnicos da bolsa de valores para classificar a direção das variações futuras, poucas utilizam a oscilação de alguns indicadores importantes para a identificação de tendência, utilizando os valores integrais ou discretos destes indicadores, o que pode torná-los menos efetivos nos algoritmos de aprendizagem de máquina. Há alguns indicadores técnicos pouco utilizados, como os indicadores direcionais. Poucos destes artigos utilizam a seleção de atributos e o estudo dos melhores parâmetros para cada indicador de análise técnica. Além disso, faltam estudos sobre o tamanho das janelas de tempo de cada indicador de análise técnica. Visto se tratar de conjuntos de dados muito ruidosos, pouco se utiliza de técnicas para amenizar este problema, como mineração de dados do tipo *ensemble*. Além disso, na validação dos resultados pouco se utiliza de algoritmos de negociação, que execute uma simulação deste mercado, tornando os números mais próximos dessa realidade. Esta modalidade pode avaliar o modelo preditivo adicionando a precisão da negociação, sendo as compras que deram retorno positivo sobre o total de compras e a taxa de retorno (lucro/prejuízo) total das negociações no período de validação.

1.2 Problema

O mercado de ações é bastante volátil para investimentos em curto prazo, e no longo prazo pode não ser rentável se o investimento ocorrer em um momento excessivamente alto ou em uma ação decadente. Sendo assim, o momento de entrada e saída de uma ação na bolsa de valores pode não representar um fator de lucro, mesmo investindo em ações com boa reputação.

A volatilidade deste mercado e suas incertezas geram conjuntos de dados bastante caóticos e a previsão do movimento do dia seguinte para este tipo de informação torna-se bastante complexa. A extração de características para este tipo de informação volta-se para os indicadores de análise técnica, normalmente utilizado por operadores do mercado financeiro. Em sua maioria, os indicadores apontam tendências, porém mesmo em movimentos tendenciosos, uma ação pode ter o número de altas bastante próximo ao número de baixas alterando apenas a escala entre as altas e baixas, conforme apresentado em um trecho do gráfico da ação Gerdal PN negociada na bolsa de valores de São Paulo entre o final de julho e início de setembro de 2008.

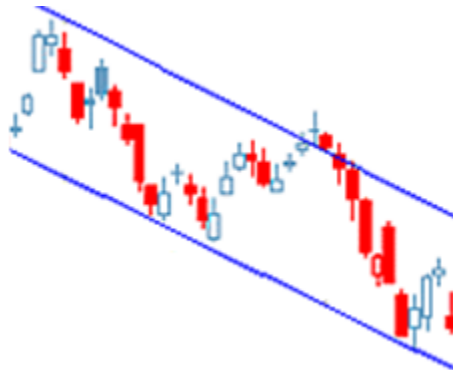


Figura 1.1: Exemplo de tendência de baixa com muitos ruídos

Ao total, o gráfico apresentado na figura 1.1 apresenta 18 dias de alta (azul) e 18 dias de baixa (vermelho) em um período visualmente com tendência de baixa conforme as linhas traçadas em azul, o que leva ao declínio no preço da ação neste caso é a escala dos movimentos dos dias de baixa que maiores comparados aos dias de alta. Esta característica da informação leva os classificadores baseados em previsão do dia seguinte a erros no aprendizado, pois a classificação é binária e discreta: subirá ou cairá, não levando

em consideração a escala do movimento. Além desta característica, uma ação nem sempre está em tendência e pode trocar de tendência repentinamente. Formando com estas características, um conjunto extremamente ruidoso e com as taxas de classificação do dia seguinte comprometidas. Desta forma, as taxas de acerto de algoritmos de classificação do movimento do dia seguinte para este tipo de informação são consideradas baixas se comparados à outras aplicações. O maior desafio de algoritmos que propõem-se a classificar o dia seguinte é sobrepor esta natureza ruidosa com técnicas de pré-processamento e otimização que possam gerar um modelo de classificação generalista, com baixa variância de resultados e que possam ser adicionadas à um algoritmo de negociação obtendo retornos relevantes.

1.3 Objetivos

O objetivo geral deste trabalho é criar um método de classificação automática do movimento do dia seguinte em ações da bolsa de valores, classificando de forma binária se o preço da ação irá se valorizar ou não no dia seguinte por meio de indicadores de análise técnica do mercado financeiro. Para sobrepor os problemas citados na seção 1.2, serão utilizados o aprendizado de máquina *Random Forest* combinado com seleção de atributos ReliefF, opções mais tolerantes a ruídos. E para validar os resultados será utilizado a taxa acerto do classificador e o lucro obtido por um algoritmo de negociação, com o intuito de avaliar o comportamento em uma situação mais próxima da realidade deste mercado. Para alcançar este propósito, serão considerados os seguintes objetivos específicos:

- Encontrar através de um algoritmo genético o tamanho de janela ótimo para cada atributo selecionado.
- Obter o melhor conjunto de atributos para a previsão de valores de cada ação através da seleção de atributos.
- Encontrar através de outro algoritmo genético o melhor formato para a discretização de cada atributo.

1.4 Hipótese

Identificar os melhores tamanhos de janelas de tempo para cada indicador técnico e seu formato de discretização através de algoritmos meta-heurísticos aumentam a precisão do classificador e

consequentemente do algoritmo de simulação de negociação que utiliza o modelo construído com os dados não otimizados.

1.5 Organização do Documento

Este trabalho está organizado em 6 capítulos. O Capítulo 2 apresenta a fundamentação teórica, onde é apresentada a introdução dos assuntos abordados, descrevendo conceitos importantes para esse trabalho. No Capítulo 3 é apresentado estado da arte do tema proposto. O Capítulo 4 apresenta o método aplicado nessa pesquisa. O Capítulo 5 descreve os resultados obtidos. No Capítulo 6 é apresentada a conclusão e trabalhos futuros. E por fim, o Apêndice contém todos os resultados dos experimentos realizados.

Capítulo 2

Fundamentação Teórica

Neste capítulo serão discutidos alguns conceitos da aplicação de mineração de dados para classificação da movimentação das ações de bolsa de valores. Dessa forma, serão apresentados temas como a previsão de tendências do mercado de ações através de análise de séries temporais. Para um melhor entendimento, o capítulo foi dividido em: “Mercado de Ações”, “Previsão em Séries Temporais” e “Previsão em Séries Temporais Financeiras”. Com a análise dos princípios técnicos e seus modelos de previsão, pode-se fundamentar uma proposta que possa melhorar os atuais modelos mais eficientes.

2.1 Mercado de Ações

Ações são títulos que representam uma fração do capital social de uma empresa. Empresas vendem uma participação societária através de ações no mercado procurando uma forma barata de obter capital e, desta forma, aplicar estes novos recursos em melhorias que possibilitem o crescimento da empresa e o aumento dos lucros. Os compradores das ações tornam-se sócios da empresa e participam dos lucros da empresa através da distribuição de dividendos. Os compradores de ações também têm a expectativa de crescimento da empresa e com isso lucrar com a valorização do valor da ação negociada na bolsa de valores.

O valor de determinadas ações varia de acordo com as leis de mercado, como a lei da oferta e da procura e também de fatores macroeconômicos, como eventos políticos, situação econômica, situação

mercadológica, taxas e juros bancários, expectativas de investidores, decisões de investidores institucionais, movimentos de outros mercados de ações e psicologia dos investidores (Miao, Chen, e Zhao 2007).

2.2 Previsão em Séries Temporais

O conjunto de dados de uma ação tem a característica de ser uma sequência de valores separados por unidades de tempo. Dentro dos estudos em mineração de dados, este tipo de conjunto é chamado de série temporal. Séries temporais podem ser obtidas facilmente de aplicações científicas e financeiras. Além da sequência de preços de uma ação, as séries temporais podem ser encontradas em eletrocardiogramas, temperaturas diárias, vendas semanais de produtos, entre outros. Os dados de series temporais podem estar em grande quantidade, alta dimensionalidade e ser atualizado continuamente. Sua natureza normalmente é numérica. A partir destes dados pode-se extrair informações do passado para estudar comportamentos futuros através da aplicação adequada de métodos de análise dos dados temporais, como por similaridade e através da comparação de padrões sequenciais (Fu 2011).

A utilização de séries temporais para cálculos, planejamento e previsões datam da década de 1950, com o trabalho de Robert G. Brown que estendeu seu trabalho de suavização exponencial simples para dados discretos e métodos desenvolvidos para extração de tendências e sazonalidade. Esse estudo foi aplicado na gestão de estoques da marinha americana, que resultou em suas publicações (R. G. Brown 1959). Posteriormente, Brown integrou sua gestão de estoque com o planejamento de produção e controle, gerindo uma cadeia de produção baseado na previsão em séries temporais (Robert Goodell Brown 1963).

Também na década de 1950, Charles C. Holt, trabalhou independentemente de Brown no braço logístico de pesquisa naval da marinha americana e publicou um método similar de suavização exponencial de tendências aditivas. Mais tarde, em 1960, Winters testou o método de Holt com dados empíricos e o método passou a ser conhecido como sistema de previsão Holt-Winters (Holt 1957). Os métodos de suavização exponencial de Holt também foram publicados no livro “Planejamento de Produção, Estoque e Força de Trabalho” em 1960, de Holt, Modigliani, Muth, e Simon. Os trabalhos de Brown, Holt e Winters foram fundamentos para os estudos em séries temporais econômicas e de negócios, que são dados com tendências, padrões sazonais e ciclos. Previsões acuradas para estes tipos de dados são vitais para decisões nos mercados de varejo, *marketing*, produção e muitos outros setores de negócios (Gan et al. 2014).

2.3 Características da Previsão em Séries Temporais Financeiras

Das pesquisas em mineração de séries temporais para previsão de valores futuros, pode-se extrair a identificação de padrões e estratificação de características temporais como sazonalidade, ciclos, tendência e caminhada aleatória. Em adição à identificação de padrões e características temporais, combinam-se técnicas de mineração de dados, como associação por regras, agrupamento, métodos de classificação, sumarização e regressão (Fu 2011). Ademais, as pesquisas de séries temporais separam-se em duas áreas: séries temporais estacionárias e não estacionárias. A diferença entre essas duas áreas é o comportamento da média e do desvio padrão sendo iguais a zero para o caso das séries estacionárias (Fu 2011).

O mercado de ações é um exemplo de series não estacionárias. Em particular, há períodos que podem conter tendências, ciclos e caminhadas aleatórias (*random walks*) ou uma combinação de mais de uma dessas três. Além disso, a previsão em séries temporais do mercado de ações é essencialmente dinâmica, não linear, complexa, não paramétrica e caótica (Abu-Mostafa e Atiya 1996).

Apesar das várias influências neste tipo de série temporal, as ações são principalmente afetadas por tendências e caminhadas aleatórias. As caminhas aleatórias, por definição, não podem ser previstas. Portanto, analistas deste mercado analisam uma combinação de fatores para que possam identificar padrões e estatísticas que indiquem o início e o fim de tendências. As tendências são formadas principalmente por fatores externos, como a situação econômica do país e global, que pode se comportar na forma de ciclos econômicos, onde o ciclo tem início, meio e fim, formando um período de uma longa tendência econômica, podendo ser subdivida em várias sub-tendências menores e períodos de estagnação. Conforme teorias dos ciclos econômicos (Kalecki 1977; Mises 1978), um alto ou baixo crescimento tem efeitos auto estimulantes, o que implica dizer que crescimento pode gerar mais crescimento e decrescimento também gera mais decrescimento. Normalmente o decrescimento é remediado com expansão monetária pelos bancos centrais, levando dinheiro de baixo custo para os mercados, inclusive para o mercado de ações (Mises 1978).

Assim, ficam evidentes períodos com uma tendência definida na economia, o que influencia diretamente os valores das ações. Analistas deste mercado normalmente tomam suas decisões baseadas nestes aspectos, tanto para a tendência da economia como para a tendência dos valores das ações. Neste modelo tradicional de análise, identifica-se se o mercado está em tendência de alta, para que o investidor possa comprar a ação e manter-se até que outras implicações afetem o preço da ação e mudem a tendência. Apesar de um modelo simples, nem sempre é fácil encontrar uma tendência dominante. O mercado pode

estar no fim de uma tendência, ou não ter uma tendência definida. Portanto, encontrar um momento inicial de tendência e o momento final ainda é uma tarefa desafiadora. Contudo, algumas ferramentas de análise técnica, baseadas em estatística podem ajudar a encontrar estes pontos e são apresentadas no decorrer deste capítulo.

2.3.1 Atributos dos Preços das Ações

Para o cálculo de indicadores técnicos, que serão apresentados na seção 2.3.2, é necessário a leitura da base de dados deste tipo de informação, que é composta por cinco atributos básicos listados abaixo.

- *Open* – Este é o preço da primeira negociação do período; se a série for diária, corresponde à primeira negociação do dia.
- *High* – É o maior valor alcançado no período. Este ponto configura o ponto de saturação da alta, em que os vendedores passaram a ser maioria na negociação, forçando os preços para baixo. Por outro ponto de vista, este é o ponto máximo que os compradores aceitam pagar.
- *Low* – Valor mais baixo do período. Inversamente ao ponto mais alto, este ponto pode configurar o ponto em que tanto compradores como vendedores acharam o preço barato e o lado comprador passou a prevalecer, caso não tenha sido o preço da última negociação.
- *Close* – É o valor da última negociação do período, sendo o atributo mais utilizado para avaliar a ação.
- *Volume* – É o total de ações negociadas durante o período. O valor total negociado é resultado da multiplicação do volume com preço médio ponderado.

Estes atributos são utilizados pelos analistas para criar indicadores que extraem características do comportamento do mercado, tendência, saturação de movimentos e padrões, indicando o sentimento e intenções dos investidores (Achelis 1995).

Na próxima seção, serão explorados indicadores calculados a partir dos atributos e que são os mais utilizados na literatura recente.

2.3.2 Indicadores Financeiros

Os indicadores são valores numéricos que descrevem algum tipo de característica da ação, indicando melhores momentos para efetuar negociações com a ação. Os indicadores contêm métricas diferentes e tem diferentes maneiras de uso, por exemplo, para identificar a tendência. Os indicadores estão divididos em tendência, volume e momentum.

Mais adiante nesta seção serão apresentados os indicadores e suas equações, que são compostas basicamente por um ou mais dos cinco atributos do conjunto de dados originais representados por:

- C_t – Preço de fechamento no dia t ;
- L_t – Preço mínimo do dia t ;
- H_t – Preço máximo do dia t ;
- V_t – Volume de ações negociadas no dia t ;
- O_t – Preço de abertura das negociações no dia t .

Além dessas variáveis, normalmente os indicadores utilizam uma janela de tempo no passado para executar seu cálculo, que é representada por n . Ainda há outras duas derivações dessas variáveis, HH_{t-n} e LL_{t-n} , preço máximo e preço mínimo, respectivamente, dentro de um período n no passado.

2.3.2.1 Indicadores Baseados em Tendência

São indicadores que extraem a tendência dos dados do mercado. Para uma tendência de alta, o preço da ação está predominantemente subindo, gerando uma tendência de alta durante um período. Por outro lado, a tendência ainda pode ser de baixa, estando em um período com mais quedas do que altas, ou ainda, indefinido se durante o período a ação obteve altas e baixas equivalentes. O mercado em alta também é conhecido como *bullish market*, e o mercado em baixa como *bearish market*.

2.3.2.2 Indicadores Momentum

São indicadores que auxiliam a identificar a tendência. Momentum são indicadores que confirmam a continuidade da tendência, ou a sua fraqueza. Devem ser utilizados com indicadores de tendência.

2.3.2.3 Indicadores Baseados em Volume

O volume negociado é um fator importante na identificação de um aumento no interesse ou desinteresse de uma ação. Estes indicadores além de indicar a continuidade de uma tendência ou sua fraqueza, podem indicar possíveis reversões de tendência. Os indicadores desta categoria utilizam a variação do preço e do volume para indicar se a mudança de interesse dos investidores é positiva ou negativa.

2.3.2.4 Indicadores Osciladores

A oscilação é um formato de indicador, sendo compatível com cálculos de tendência, momentum, volatilidade e volume. São considerados osciladores todos os indicadores que variam em uma faixa de valores ou sobre um eixo. Pode-se transformar um indicador em oscilador, como a diferença entre duas médias ou variação do preço da ação nos últimos dias. Há também, indicadores que são calculados para descobrir pontos de sobrecompra e sobrevenda (John J Murphy 1999).

2.3.2.5 *Moving Average* (MA)

Média móvel (*moving average* - MA) é um modelo simples que tenta prever valores futuros através da média de uma sequência de valores de subconjuntos da série temporal em análise (Bollerslev e Domowitz 1993). A média móvel é um dos métodos de suavização de séries temporais econômicas, a qual retira parte dos caminhos aleatórios, tornando possível indicar a tendência de uma ação.

Este indicador é definido por Bollerslev e Domowitz (1993) conforme a equação (2.1), na qual x_t é a série de previsão, β_i são os pesos para cada deslocamento de tempo, e μ_t é uma sequência com médias calculadas da série original. O modelo utiliza as últimas n médias, e quanto menor é seu valor, mais rápido os valores passados são esquecidos pelo modelo.

$$x_t = \beta_0 + \mu_t + \sum_{i=1}^n \beta_i \mu_{t-i} \quad (2.1)$$

A média móvel simples (*Simple Moving Average* - SMA) é expressa sem os pesos, aplicando a média para todos os valores anteriores e o atual, sendo expresso como indicado na equação 2.2:

$$SMA_t = \sum_{i=0}^n \mu_{t-i} \quad (2.2)$$

Ou também, como descrito por Kim 2003, indicado através da equação 2.3, onde C_t seria o preço de fechamento da ação e n o tamanho da janela de tempo a ser considerada no passado para o cálculo.

$$SMA_t = \frac{\sum_{i=0}^n C_{t-i}}{n} \quad (2.3)$$

A partir deste modelo, criou-se algumas derivações. Dentre elas, a *Exponential Moving Average* (EMA), conhecido também como *Exponentially Weighted Moving Average* (EWMA), definido por Roberts (1959), que dá pesos maiores para β para períodos mais recentes. Os pesos são definidos por: $\beta = e^{-\alpha}$.

Outra derivação de médias móveis que dá mais importância aos movimentos recentes é a *Weighted Moving Average* (WMA). Semelhante à proposta de Roberts, no WMA os pesos não são distribuídos de forma exponencial e sim pelo produto de n (Kim 2003), conforme equação 2.4.

$$WMA_t = \frac{((n) \times C_t + (n-1) \times C_{t-1} + \dots + C_{10})}{(n + (n-1) + \dots + 1)} \quad (2.4)$$

2.3.2.6 *Commodity Channel Index* (CCI)

Introduzido por Donald Lambert em 1980, este oscilador de momentum mede a variação do seu preço com sua média. Inicialmente utilizado para identificar ciclos nos preços de commodities, como o petróleo, logo foi utilizado para câmbio e ações (John J Murphy 1999).

Em sua definição, Lambert atribuiu uma constante no valor de 0,015 para a multiplicação com o desvio padrão. Esta constante assegura que cerca de 70% a 80% dos valores ficam entre -100 e +100. A fórmula está definida na equação 2.5, sendo C_t o último preço de fechamento, SMA_t o valor da média móvel simples atual e σ_t o desvio padrão também para o período atual:

$$CCI_t = \frac{C_t - SMA_t}{0,015\sigma_t} \quad (2.5)$$

Para o entendimento do representa este indicador, é necessário entender os conceitos de sobrecomprado e sobrevendido. Sobrecomprado, neste contexto, entende-se por um período dias subsequentes em que houve uma alta expressiva sem haver baixas relevantes para amenizar a alta e permitir novos compradores, e sobrevendido seria o inverso, uma ação sobrevendida sofreu quedas subsequentes sem haver altas ou com poucas altas. Uma ação sobrecomprada ou sobrevendida pode indicar que o movimento está se esgotando. A interpretação dos resultados do indicador CCI utiliza-se deste conceito, pode-se entender que o mercado está sobrecomprado quando está acima de +100, e sobrevendido quando está abaixo de -100. Contudo, 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.

2.3.2.7 *Relative Strength Index* (RSI)

Também conhecido como índice de força relativa, foi desenvolvido por J. Welles Wilder. É um oscilador cujos valores resultantes ficam entre 0 e 100 e seu objetivo é identificar se o mercado está sobrecomprado (valores próximos de 100) ou sobrevendido (valores próximos a 0). Sua formulação é baseada apenas na movimentação dos preços, e seus valores sobem quando a variação de preços nos dias de alta é mais expressiva e há uma variação para baixo nos dias de baixa, conforme formulado por Wilder, e

apresentado na equação 2.6, sendo Up a variação dos dias positivos, Dw a variação dos dias negativos e n o tamanho da janela de tempo no passado que o indicador deve levar em consideração. Wilder ainda sugeriu que o período amostral seja de 14 períodos, 14 dias em uma série temporal diária.

$$RSI_t = 100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i}) / (\sum_{i=0}^{n-1} Dw_{t-i}/n)} \quad (2.6)$$

Entretanto, a utilização do RSI não é interessante para identificação de tendências, pois em uma tendência o RSI pode ficar próximo de uma das margens por várias vezes, sem inverter a tendência. Sendo assim, não é aconselhada a utilização deste oscilador de forma isolada.

2.3.2.8 Oscilador Estocástico

Este indicador se utiliza de uma composição de duas linhas, a %K e %D. A linha principal é a %K, e a %D é a média móvel de %K. O indicador %K é calculado através da oscilação de preços em um dado período com o preço final registrado no período.

A equação 2.7 define o cálculo de %K, considerando que C_t é o preço de fechamento, LL_{t-n} é o valor mínimo alcançado no período estipulado e HH_{t-n} é o valor máximo alcançado neste mesmo período (Kim 2003).

$$\%K_t = \frac{C_t + LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100 \quad (2.7)$$

O cálculo de %D, dado pela equação 2.8, é o cálculo da média móvel de %K, assim como a média móvel é passado por parâmetro n para o número de períodos no passado a ser considerado, além de utilizar o valor do estocástico %K. Normalmente a média móvel é do tipo simples, mas em alguns casos pode-se calcular outros tipos de médias móveis como exponencial e ponderada (Achelis 1995).

$$\%D_t = \frac{1}{n} \sum_{i=0}^{n-1} \%K_{t-i} \quad (2.8)$$

O resultado sempre varia entre 0% e 100%. Sua interpretação é parecida com a relação entre preços e a média móvel. Neste caso, se o valor de %K estiver acima da média móvel (%D) está em uma tendência altista, caso contrário, é uma tendência de baixa (John J Murphy 1999).

2.3.2.9 *Moving Average Convergence/Divergence* (MACD)

Proposto por Gerald Appel, o MACD é um indicador de momentum baseado em tendência. A formulação do MACD é a diferença entre a média móvel exponencial (EMA) de longo período, que por padrão é 26 dias, e a média móvel exponencial curta, que por padrão é de 12 dias. A utilização do MACD normalmente ocorre pelo cruzamento das médias móveis exponenciais, conforme demonstrado na Figura 2.1 gerado pelo aplicativo TraderZone; o cruzamento de baixo para cima da média de 12 dias indica a compra, e para baixo indica a venda (John J Murphy 1999; Abu-Mostafa e Atiya 1996).

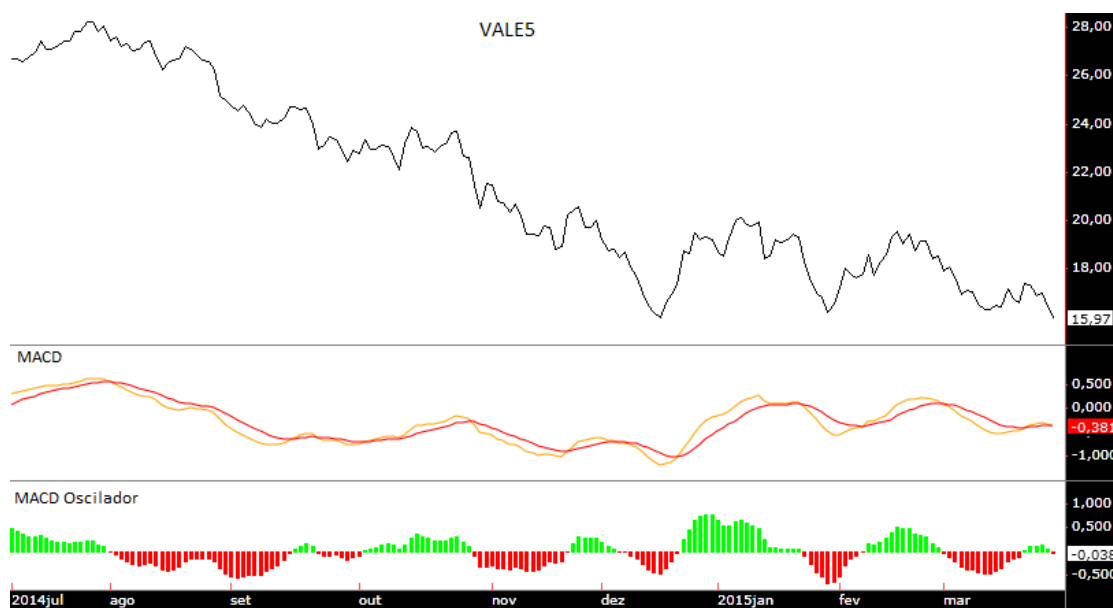


Figura 2.1: MACD e Oscilador MACD

Na parte superior da figura 2.1, é apresentado a série temporal da evolução do preço da ação VALE5 negociada na bolsa de valores de São Paulo. Logo abaixo, é apresentado as duas médias móveis exponenciais, de 26 dias em vermelho e 12 dias em amarelo. Na parte inferior é apresentada a oscilação

entre essas duas linhas, ou seja, a diferença entre a média de 12 e a de 26. Resultando em valores positivos quando o mercado está em tendência de alta e valores negativos quando o contrário ocorre.

2.3.2.10 *Accumulation/Distribution (A/D)*

Este é um indicador que associa as movimentações de preços com o volume negociado. Este indicador pode revelar a importância dos movimentos de preços, baseado na premissa de que quanto maior o volume negociado, maior a importância do movimento (Achelis 1995). Quando há uma movimentação para cima neste indicador, significa que há uma acumulação de ações em posse de investidores que compraram recentemente. Quando este indicador reduz seu valor, é interpretado que há uma distribuição de ações no mercado, ou seja, os investidores que acumularam estas ações estão vendendo seus papéis, sendo calculado de acordo com a equação 2.9, sendo C_t o valor de fechamento do dia, L_t o valor mínimo alcançado no dia, H_t o valor máximo alcançado no dia e V_t o volume total do dia.

$$AD_t = \frac{[(C_t - L_t) - (H_t - C_t)]}{(H_t - L_t) \times V_t} \quad (2.9)$$

2.3.2.11 Williams %R

Desenvolvido por Larry Williams, este é outro indicador de momentum, e calcula níveis de sobrecompra e de sobrevenda. Seu indicador é apresentado normalmente entre 0% e 100%, e são considerados níveis de sobrecompra os patamares acima de 80% e sobrevenda aqueles abaixo de 20%.

Semelhante ao MACD, seu cálculo utiliza os valores máximos e mínimos em dado período, conforme a equação 2.10, sendo HH_{t-n} o valor mais alto alcançado no período de n dias anteriores, C_t o valor de fechamento da data atual e L_t o valor mínimo da data atual.

$$WR_t = \frac{HH_{t-n} - C_t}{HH_{t-n} - L_t} \times 100 \quad (2.10)$$

2.3.2.12 Psychological Line (PSY)

Indicador no formato de oscilador, é a razão do número de períodos em alta sobre o total de períodos, indica a força da atual tendência, seus valores podem ser vistos como sobre compra ou sobre venda da ação. O cálculo é demonstrado na equação 2.11, sendo NUp o número de altas no período n .

$$PSY_t = \frac{\sum_{i=0}^n NUp_{t-i}}{n} \times 100 \quad (2.11)$$

2.3.2.13 Indicadores Direcionais

Os indicadores direcionais, compostos por *Plus Directional Indicator* (+DI), *Minus Directional Indicator* (-DI) e *Average Directional Index* (ADX), criados por Wilder Jr, (1978), são indicadores que auxiliam a identificar a direção e força da tendência.

Para o cálculo desses indicadores é necessário calcular o *Plus Directional Movement* (+DM) e o *Minus Directional Movement* (-DM), como indicado nas equações 2.12 e 2.13, e o *True Range* (TR), indicado na equação 2.14, sendo C_t o valor de fechamento do dia, L_t o valor mínimo alcançado no dia, H_t o valor máximo alcançado no dia.

$$+DM_t = \begin{cases} H_t - H_{t-1} & \forall H_t - H_{t-1} > 0 \\ 0 & \text{se não} \end{cases} \quad (2.12)$$

$$-DM_t = \begin{cases} L_{t-1} - L_t & \forall L_{t-1} - L_t > 0 \\ 0 & \text{se não} \end{cases} \quad (2.13)$$

$$TR_t = \max(H_t - L_t, |H_t - C_{t-1}|, |L_t - C_{t-1}|) \quad (2.14)$$

Para o cálculo de +DI é utilizado a somatória de +DM por n períodos e dividido por TR, conforme indicado na equação 2.15.

$$+DI_t = \frac{\sum_{i=0}^n +DM_{t-i}}{\sum_i TR_{t-i}} \times 100 \quad (2.15)$$

Porém, para o cálculo do indicador -DI utiliza-se o -DM, conforme a equação 2.16.

$$-DI_t = \frac{\sum_{i=0}^n -DM_{t-i}}{\sum_i TR_{t-i}} \times 100 \quad (2.16)$$

Os indicadores +DI e -DI indicam a direção da tendência, que é de alta se +DI for maior que -DI, e de baixa se -DI for maior que +DI. Por fim, Wilder Jr (1978) ainda desenvolveu o indicador ADX, para ser utilizado em conjunto com +DI e -DI. O ADX, é o indicador que indica a força da tendência, seja ela de alta ou de baixa, e pode ser calculado através da equação 2.17.

$$ADX_t = \sum_{i=0}^n \left(\frac{|+DI_{t-i} - (-DI_{t-i})|}{+DI_{t-i} + (-DI_{t-i})} \right) \times 100/n \quad (2.17)$$

Na figura 2.2 é apresentado um gráfico gerado pelo aplicativo TraderZone mostrando o comportamento destes três indicadores. Antes do ponto 1 e após o ponto 2 representados no gráfico com uma linha pontilhada, há uma tendência indefinida, pois os valores de -DI e +DI estão bastante próximos e pouco definidos, porém, no centro entre os pontos 1 e 2, há uma tendência negativa apontada pelos indicadores com a linha -DI superiores à linha +DI.



Figura 2.2: Indicadores de Direcionais, ação Petrobrás (PETR4) de 09/14 - 02/15

2.3.2.14 Indicadores Aroon

Indicadores Aroon são compostos por dois indicadores, *Aroon Up* e *Aroon Down*. São indicadores de momentum em formato de oscilador flutuando entre 0 e 100. Similar aos indicadores de direção, estes indicadores determinam quando uma ação está em tendência, qual a direção e quão forte. O indicador *Aroon Up* é calculado utilizando o número de dias desde de a última cotação máxima dentro de um período n , dividido por n , conforme indicado na equação 2.18, sendo n_{LastHH} o número de dias desde o último valor máximo.

$$AroonUp_t = \left[\frac{(n - n_{LastHH})}{n} \right] \times 100 \quad (2.18)$$

Por outro lado, o *Aroon Down* é calculado utilizando o número de dias desde a última cotação mínima, conforme indicado na equação 2.19, sendo n_{LastLL} o último valor mínimo.

$$AroonDown_t = \left[\frac{(n - n_{LastLL})}{n} \right] \times 100 \quad (2.19)$$

Como demonstrado na figura 2.3, o indicador *Aroon Down* está predominantemente acima do indicador *Aroon Up*, indicando uma tendência de baixa para a ação neste período.



Figura 2.3: Indicadores Aroon, ação Petrobrás (PETR4) de 09/14 - 02/15

2.3.2.15 Transformação de Indicadores em Viés e Osciladores

Os indicadores têm leituras diferentes, e são compreendidos de maneiras diferentes pelos analistas. Os indicadores no formato de oscilador já extraem valores que podem ser lidos em qualquer parte da série temporal como valores indicados para compra ou não de uma ação. Porém, alguns indicadores acompanham alguns dados originais da base de dados, como o preço de fechamento, e por isso os mesmos valores não representam sinais de compra ou venda em toda a série temporal. Um exemplo disso é o SMA, que segue o preço da ação, que pode variar muito durante os anos. Para resolver este problema, alguns autores utilizaram apenas o viés do indicador, descrito por Kim (2003) como disparidade, e como viés por Chang, Fan e Lin

(2011). O cálculo do viés proposto por Chang, Fan e Lin (2011) é apresentado na equação 2.20, como exemplo do viés entre o preço da ação e a SMA:

$$Viés_t = \frac{C_t - SMA_t}{SMA_t} \quad (2.20)$$

Além do viés, também é possível calcular a variação do indicador no período n, como indicado na equação 2.21. O indicador resultante será um oscilador.

$$\Delta Indicator_t = Indicator_t - Indicator_{t-n} \quad (2.21)$$

Com estas transformações para os indicadores não osciladores será possível extrair melhor as características para todo o período da série temporal durante o processo de mineração de dados.

2.4 Descoberta do Conhecimento em Séries Temporais Financeiras

As técnicas de mineração de dados utilizadas na previsão de valores no mercado financeiro são regras de associação, agrupamento, classificação, sumarização e regressão (Fu 2011). E para alcançar os objetivos deste trabalho, serão utilizadas técnicas de classificação pois pretende-se indicar no momento atual se o mercado é favorável à compra ou venda. Contudo, os autores normalmente utilizam a mineração de dados dentro do processo de Descoberta do Conhecimento em Base de Dados (do inglês *Knowledge Discovery in Database* - KDD), adicionando algumas técnicas como transformação de atributos em indicadores técnicos, seleção de atributos, discretização, e em alguns casos combinando o KDD com algoritmos meta-heurísticos para otimização dos dados de entrada e de parâmetros do algoritmo de classificação.

2.4.1 Métodos de Classificação

As técnicas de classificação mais utilizadas na literatura recente são RNA e SVM. *Random Forest* foi testado por Jigar Patel (2015), obtendo resultados bem-sucedidos para modelar e prever séries temporais financeiras.

Redes Neurais Artificiais

Redes neurais artificiais foram concebidas a partir do modelo cerebral humano. São modelos compostos por neurônios interligados entre camadas, de acordo com o tipo de utilização e quantidade de variáveis de entrada e saída. Os neurônios são unidades fundamentais de processamento, sendo que cada neurônio tem terminais de entrada de valores, sejam valores de entrada iniciais do modelo ou da camada anterior. Os valores recebidos são ponderados e combinados e o valor de saída do neurônio é a resposta para a sua entrada de valores. Esta resposta é encaminhada para a próxima camada ou para a saída. Aos valores recebidos pelos neurônios são atribuídos pesos. O poder das redes neurais está em capturar relacionamentos indiretos entre as variáveis (Faceli 2011). Contudo, muitos estudos apontaram que RNAs apresentam algumas limitações no aprendizado de padrões neste tipo de séries temporais. Dados com grande ruído, não estacionários e dimensões complexas são problemáticos para as RNAs (Kim 2003).

Support Vector Machines

Outro importante algoritmo de classificação, o SVM (Máquinas de Vetores Suporte, do inglês *Support Vector Machines*), desenvolvido por Vapnik e seus colaboradores (Vapnik 1998), é um método baseado na Minimização de Risco Estrutural (MRS - do inglês *Structural Risk Minimization*), utilizando-se de conceitos de modelos lineares para encontrar os limites em atributos lineares e não lineares, utilizando dimensões adicionais. Neste espaço com novas dimensões é definido um hiperplano com a melhor separação entre as instâncias, baseando-se no fato de que esta separação ótima no espaço dimensional superior corresponde à separação ótima na dimensão original (Kim 2003). O processo de treinamento do SVM é o equivalente a resolver um problema de programação quadrática linearmente restrita. Sendo o objetivo do SVM encontrar uma regra de decisão com a maior generalização possível e o menor erro dentro do próprio conjunto de treinamento. Ao contrário de muitos algoritmos de aprendizado de máquina, o SVM obtém uma

solução única e ótima global. Por outro lado, o SVM pode exigir excessivamente os recursos de computação para processar bases com grande quantidade de registros (Ratnadip Adhikari e R. K. Agrawal 2013).

Random Forest

Por fim, o algoritmo testado por Jigar Patel (2015) e que obteve resultados relevantes, *Random Forest*, é um algoritmo de aprendizado do tipo *ensemble* que utiliza árvores de decisão. Algoritmos de aprendizado de máquina do tipo *ensemble* criam um classificador no estágio de treinamento agrupando vários classificadores. Uma árvore de decisão é um modelo representado por uma árvore com decisões nos nós e as classes nas folhas (Witten, Frank, e Hall 2011). Os algoritmos de árvore de decisão mais conhecidos são CART (L Breiman et al. 1984), ID3 (Quinlan 1986) e C4.5 (Quinlan 1993). Porém, árvores de decisão costumam resultar em um classificador com sobre ajuste com o conjunto de dados de treinamento, principalmente causado pela característica de haver um alto nível de viés e variância no seu aprendizado. Contudo, estes problemas podem ser contornados com o uso de algoritmos do tipo *ensemble*, como *boosting* ou *bagging*. Estes algoritmos criam m classificadores e selecionam do conjunto de treinamento apenas uma amostra dos dados para a construção dos classificadores. Na fase de classificação, o algoritmo *ensemble* testa as instâncias de teste e determina a classe para cada instância através da maioria dos votos. Como demonstrado por Leo Breiman (2001) este processo reduz o viés e a variância dos resultados.

Leo Breiman (1994) propôs o algoritmo *Bagging*, que é um algoritmo de aprendizado de máquina do tipo *ensemble*, que posteriormente foi utilizado para compor o *Random Forest*. *Bagging* gera m conjuntos de treinamento, capturando amostras dos dados originais uniformemente e com reposição. Porém, no estudo feito por Leo Breiman (2001) o algoritmo de *Bagging* por si só não superou outros algoritmos de *Boosting*. Contudo, Leo Breiman (2001) alcançou em seus estudos resultados mais consistentes escolhendo o algoritmo de *Bagging* acrescentando a aleatoriedade de escolha do conjunto de atributos de cada árvore. Apesar deste estudo de Leo Breiman, a combinação de algoritmos do tipo *ensemble* com a aleatoriedade do conjunto de atributos de cada árvore foi introduzido antes por Tin Kam Ho (1995), que introduziu o termo "*random forest*" e Amit e Geman (1997) também publicou outras formas para o *Random Forest*. Porém, com o estudo de Leo Breiman estabeleceu-se o método de *Bagging* com aleatoriedade dos atributos para o *Random Forest*. Em outras palavras, *Random Forest* cria um classificador *Bagging* com um conjunto de árvores de decisão e selecionando aleatoriamente um conjunto de amostra com reposição de atributos para

a construção das árvores. Em um estudo de Trevor Hastie, Tibshirani e Friedman (2008), *Random Forest* reduziu o sobre ajuste de ruídos do conjuntos de treinamento por suavizar os ruídos e reduzir a variância.

Nesta área há ainda estudos utilizando classificadores baseados em árvores como ID3 e C4.5 e classificadores probabilísticos, como *Naïve-Bayes*. Contudo, a utilização de algoritmos de classificação têm obtido melhores resultados nesta área quando acompanhadas por outros métodos de pré-processamento e otimização, que serão vistos nas próximas seções.

2.4.2 Técnicas de Redução de Dimensionalidade

As técnicas de redução de dimensionalidade são comumente utilizadas na fase de pré-processamento para remover atributos indesejados. Isso se faz necessário pois os conjuntos de dados podem conter atributos que são redundantes ou irrelevantes, o que afeta negativamente os resultados das técnicas de aprendizado de máquina. Para contornar este problema, as técnicas de redução de dimensionalidade reduzem a quantidade de características do conjunto de dados original. Há duas maneiras para realizar esta redução: seleção de atributos e aprendizado de características (ou extração de características) (Van Der Maaten, Postma, e Van Den Herik 2009; Fang et al. 2014). Este trabalho se concentra na utilização de seleção de atributos.

2.4.2.1 Seleção de Atributos

Este é um processo de remoção de alguns atributos do conjunto de dados original, usando no processo de classificação apenas o melhor subconjunto de atributos encontrado pelo algoritmo. Este processo normalmente é utilizado para aumentar a assertividade do classificador. Também é utilizado para reduzir o custo computacional pela redução da dimensionalidade do conjunto de dados.

As técnicas mais usadas para reduzir o número de atributos incluem técnicas de filtro, que realiza a escolha do subconjunto de características utilizando a qualidade dos atributos independente do algoritmo de classificação, e *Wrapper*, que se utiliza de um classificador para testar qual subconjunto é o mais eficiente, normalmente o mesmo classificador que será utilizado na fase de classificação. *Wrapper* é uma técnica eficiente, porém computacionalmente muito cara.

Normalmente, os algoritmos de seleção de atributos utilizam um método para a geração do subconjunto, um método para a geração de subconjuntos sucessores e um método de avaliação do subconjunto encontrado. O processo é executado até um critério de parada. As técnicas de busca do subconjunto utilizam-se de uma estratégia para iniciar o conjunto a ser avaliado, que podem ser:

- Para frente – Inicia-se com um conjunto vazio e progressivamente adiciona atributos;
- Para trás – Inicia-se com o conjunto completo de atributos e remove sucessivamente;
- Bidirecional – Pode executar a busca para ambos os lados;

Após obter o subconjunto inicial, os algoritmos implementam a estratégia de busca. Uma das opções é a busca exaustiva do melhor subconjunto, porém, para um conjunto de n atributos a complexidade da busca será de $O(2^n)$. O que na maioria dos casos é uma complexidade computacionalmente muito alta. Para isso, diferentes heurísticas podem ser aplicadas para efetuar a busca do subconjunto. Comumente são utilizadas buscas do tipo exponencial, sequencial ou aleatória. As buscas aleatórias podem iniciar com um conjunto aleatório e na sequência seguir de maneira sequencial ou ainda seguir em todas as seleções de forma aleatória. A aleatoriedade ajuda o algoritmo a escapar da armadilha de ótimos locais (Molina, Belanche, e Nebot 2002). Além disso, as buscas são compostas por um operador para a geração do estado sucessor para adicionar ou remover atributos nos subconjuntos gerados. Pode-se utilizar da mesma estratégia da inicialização subconjunto com operadores “para frente”, “para trás”. Além do operador composto que aplica sucessivos passos para frente e sucessivos passos para trás. Outro operador possível é o ponderado, que utiliza todos os atributos, apenas utiliza pesos conforme avança o seu processo. Por último, um operador aleatório pode ser utilizado. Outro componente importante dos algoritmos de seleção de atributos é o método de avaliação dos subconjuntos, destacam-se Medidas de Distância, de Informação, de Dependência e de Consistência.

- Medidas de Distância – são cálculos computam a divergência das informações. Os atributos que resultam em maior distância entre as instâncias de classes diferentes, são atributos com maior qualidade. A distância Euclidiana e a distância de Manhattan são as mais conhecidas (Molina, Belanche, e Nebot 2002).
- Medidas de Informação – normalmente determinam o ganho de informação de um atributo. As medidas podem ser computadas através de entropia, por exemplo (Molina, Belanche, e Nebot 2002).

- Medidas de Dependência – são conhecidas também como medidas de correlação ou similaridade. As medidas de dependência avaliam a correlação com outros atributos. A premissa deste método é que bons subconjuntos são compostos de atributos com baixa correlação entre si, evitando redundância entre atributos. O algoritmo mais conhecido é o CFS (*Correlation-based Feature Selection*), que também avalia a qualidade de um atributo em classificar uma classe individualmente (Hall 1999).
- Medidas de Consistência – Outra forma de medida que busca na interação entre atributos para formar o subconjunto. Este tipo de medida, diferentemente do anterior, analisa a capacidade de a relação entre atributos prever a classe da instância (Liu e Setiono, 1996).

Recursive Elimination of Features (Relief)

Kira e Rendell (1992) propuseram o uso de métodos baseados em instâncias para selecionar atributos, chamando de *Recursive Elimination of Features* (Relief). Este algoritmo estima a qualidade de cada atributo de acordo com quão bem os seus valores são diferenciados entre as instâncias medidos pela distância entre as k classes corretas mais próximas em oposição a distância de k classes incorretas mais próximas. O algoritmo usa a geração de subconjuntos de forma aleatória, porém para a geração de sucessores os autores escolheram a utilização de um método ponderado, o qual mantém presente todos os atributos com pesos atribuídos. Contudo, o algoritmo básico Relief é bastante simples, trabalhando somente com classificação binária. Então Kononenko (1994) estendeu o algoritmo para lidar com múltiplas classes e dados com ruídos, chamando de ReliefF. Como demonstrado na figura 2.4, ReliefF suaviza os dados com ruídos pela contribuição média dos acertos mais próximos de mesma classe e os erros mais próximos. A entrada de dados do algoritmo é o conjunto de instâncias de treinamento com apenas o atributo em avaliação e outro conjunto com apenas a classe, sendo o algoritmo é testado para todos os atributos. A saída de cada atributo é a estimativa de qualidade para cada atributo.

```

1. set all weights  $W[A] := 0.0$ ;
2. for  $i := 1$  to  $m$  do begin
3. randomly select an instance  $R_i$ ;
4. find  $k$  nearest hit  $H_j$ ;
5. foreach class  $C \neq class(R_i)$  do
6. from class  $C$  find  $k$  nearest misses  $M_j(C)$ ;
7. for  $A := 1$  to  $a$  do

8.  $W[A] := W[A] - \sum_{j=1}^k diff(A, R_i, H_j) / (m \times k) + \sum_{C \neq class(R_i)} \frac{\left[ \frac{P(C)}{1 - P(class(R_i))} \sum_{j=1}^k diff(A, R_i, M_j(C)) \right]}{(m \times k)}$ ;

9. end;
```

Figura 2.4: Pseudo código do algoritmo ReliefF (Kononenko 1994)

Na figura 2.4, H é o acerto mais próximo, ou a instância mais próxima com a mesma classe, e M é a instância mais próxima com classe diferente. A equação para atualizar a qualidade da estimativa para todos os atributos $W[A]$ é apresentada na linha 8. A função $diff(A, R_i, H_j)$ e $diff(A, R_i, M_j(C))$ calcula a diferença entre os valores do atributo A e o aleatório selecionado da instancia R_i para os acertos e erros mais próximos (Robnik-Šikonja e Kononenko 2003). Em estudos recentes, o ReliefF mostra boa performance comparado com os demais quando os dados são contínuos e com muitos ruídos. É robusto também para interações entre atributos, porém pode ser danoso para sua comparação se houver atributos redundantes ou poucas instâncias de treinamento.

2.4.3 Técnicas de Discretização

Discretização é um processo de transformar atributos numéricos em atributos categóricos. Este processo é normalmente utilizado para obter melhores resultados na fase de classificação melhorando o desempenho do classificador em alguns casos.

Há várias técnicas de discretização, sendo que elas estão subdivididas em discretização supervisionada e não supervisionada. As técnicas supervisionadas analisam a classe da instância no processo de definição da margem de cada categoria. Porém estas técnicas podem não apresentar ganhos em dados com muitos

ruídos. Por outro lado, as técnicas não supervisionadas não levam em conta a classe da instância (Witten, Frank, e Hall 2011).

Este trabalho irá utilizar duas técnicas de discretização não supervisionadas, *Equal Width Discretization* (EWD) e *Equal Frequency Discretization* (EFD). A técnica EWD divide o intervalo de valores dos atributos em tamanhos iguais. No entanto, esta distribuição de categorias pode ser muito desbalanceada e às vezes pode dificultar a capacidade de o atributo ajudar na construção do modelo de decisão. Comumente o atributo tem algumas categorias que contêm muitas instâncias, enquanto outros contêm nenhuma. Ao contrário do EWD, o EFD desloca as margens dos valores para que todas as categorias tenham o mesmo número de instâncias de treinamento, em outras palavras, os limites das categorias são definidos pela distribuição, sendo que em cada categoria há o mesmo número de instâncias.

2.4.4 Algoritmos Genéticos

Algoritmo Genético (AG) é um tipo de algoritmo de busca e otimização meta-heurístico inspirado pela seleção natural. Ele foi concebido por John Holland no final dos anos 60 e é utilizado como um algoritmo de busca para encontrar a melhor hipótese dentro de um espaço de hipóteses candidatas, chamado de população. O AG testa a cada rodada evolutiva um conjunto de cromossomos ou hipóteses candidatas, como parâmetro de uma função de maximização ou minimização chamada função *fitness*. A função *fitness* é específica para cada aplicação. Em seu processo o AG efetua a atualização da população, através do cruzamento, mutação e seleção dos indivíduos, mantendo probabilisticamente os melhores resultados para o cruzamento, e descartando os demais. O algoritmo básico para o AG é detalhado na figura 2.5, sendo *Fitness* a função que calcula o valor dada a hipótese, p o número de hipóteses para ser incluída na população, r a fração da população a ser substituída a cada rodada pelo cruzamento dos indivíduos, m a taxa de mutação e h a hipótese em avaliação (T. Mitchell 1990).

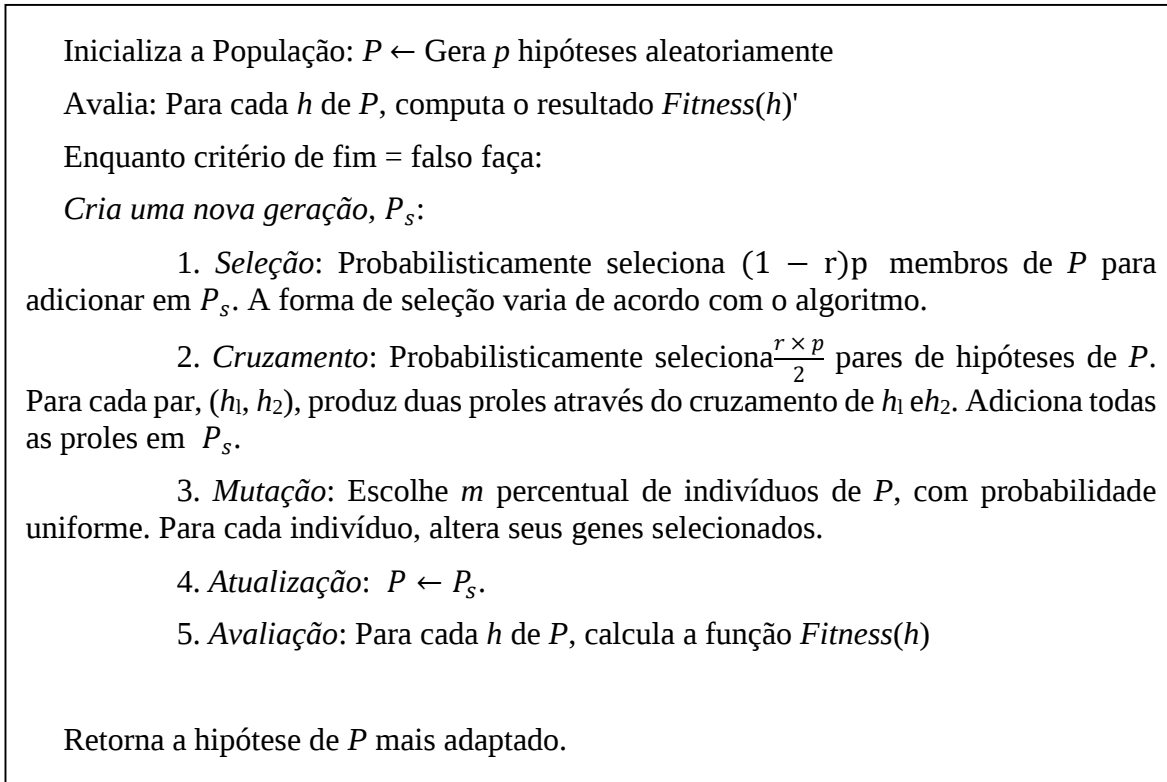


Figura 2.5: Algoritmo genético básico

As variações dos algoritmos evolutivos encontram-se exatamente na seleção, cruzamento e mutação dos indivíduos (Goldberg e Deb 1991). No trabalho original de Holland, foi utilizado o algoritmo *Fitness Proportionate Selection*, no qual a probabilidade de um indivíduo ser selecionado para reprodução é dada proporcionalmente ao resultado de sua função *Fitness*, conforme apresentado na equação 2.22. Este tipo de abordagem leva, muitas vezes, a uma convergência prematura (M. Mitchell 1998).

$$Pr(h_i) = \frac{Fitness(h_i)}{\sum_{j=0}^p Fitness(h_j)} \quad (2.22)$$

Para resolver este tipo de problema, alguns outros algoritmos foram propostos, como *Sigma Scaling*, *Elitism*, *Rank* e *Steady-State Selection* (M. Mitchell 1998). No contexto deste trabalho, o algoritmo AG utilizado é o *Elitism Selection*, no qual seleciona os mais bem-sucedidos para a próxima geração, diminuindo

a probabilidade de que indivíduos com baixo resultado na função *Fitness* se reproduzam e sejam mantidos para a geração seguinte.

Neste trabalho, o AG é utilizado para maximizar a precisão dos modelos de classificação, podendo-se utilizar AG para descobrir os melhores parâmetros ou variáveis de entrada no pré-processamento ou classificação. Kim (2000) propôs um algoritmo genético para discretização, *Genetic Algorithm Feature Discretization* (GAFD); seu algoritmo procura os melhores limites para intervalos das categorias de cada atributo além de otimizar os pesos das conexões da rede neural artificial RNA. Em um trabalho semelhante, Debie, Shafi, Merrick e Lokan (2014) utilizaram AG em uma abordagem de deslocamento de janela incremental para encontrar os limiares para a discretização de atributos contínuos. Em outro caso, Chang et al. (2011) utilizaram AG para encontrar *clusters* de ações através de semelhança entre os movimentos de preços de ações.

2.5 Considerações

É importante a construção de um modelo que otimize a fase de transformação dos dados, pois esta fase é crucial para o algoritmo indicar o movimento do próximo dia. A maioria dos indicadores de análise técnica recebem como parâmetro um tamanho de janela para se levar em consideração no cálculo, neste capítulo indicado como variável n nas fórmulas. A variação destes valores afeta o resultado, o que pode reduzir a precisão do algoritmo em indicar o movimento do próximo dia. Para isso pode-se buscar qual valor de n que gera o melhor resultado na previsão para cada indicador de análise técnica. Além disso, a utilização de todos os indicadores pode resultar em dados de alta dimensão e ocasionar um aumento na entropia, para isso a seleção de atributos pode reduzir a quantidade de atributos. Também pode-se estudar o impacto de algoritmos de discretização, para isso, o algoritmo deve capturar a quantidade de categorias e a forma de dividir estas categorias que resulte em uma maior precisão do classificador.

Capítulo 3

Estado da Arte

Neste capítulo são apresentados os trabalhos mais recentes e seus resultados sobre o tema abordado. Com a análise dos princípios técnicos, seus modelos de previsão e os resultados mais promissores da literatura recente, pode-se fundamentar uma proposta que possa melhorar os atuais modelos mais eficientes.

Kim (2003) utilizou SVM para classificar a direção do movimento das ações e em seu estudo comparou com RNA utilizando *back-propagation* (BPN) e *case-based reasoning* (CBR). Como entrada de dados para os modelos, ele utilizou 12 indicadores de análise técnica, transformados em valores discretos através de algoritmos genéticos para otimizar os valores, algoritmo já aplicado em seu estudo em (Kim e Han 2000). Os indicadores utilizados por Kim são estocásticos: K%, estocástico D%, estocástico lento D%, momentum, ROC, Williams %R, oscilador A/D, disparidade5, disparidade10, OSCP, CCI e RSI. Ele utilizou o sistema para classificação valores diários discretos em "0" para os dias que o preço da ação caiu e "1" para os dias que a ação subiu. Além disso, utilizou escalas lineares para normalizar os valores entre -1 e 1 para os dados originais, evitando que valores menores sobrepusessem valores maiores de períodos diferentes no aprendizado do SVM. Em seu estudo, Kim encontrou valores superiores utilizando SVM em relação ao RNA BPN e CBR.

Mais tarde, Kara, Acar Boyacioglu e Baykan (2011) estenderam o trabalho de Kim, também utilizando como entrada de valores para seu modelo dados diários apenas como valores discretos, indicando se o preço da ação subiu ou caiu. Porém, eles utilizaram apenas 10 indicadores técnicos dos 12 utilizados e estudados por Kim (2003) como entrada de valores para o treinamento do modelo, e com alterações nos

parâmetros dos algoritmos de classificação. Os indicadores foram SMA de 10 dias, WMA de 10 dias, *Momentum*, estocástico K%, estocástico D%, RSI, MACD, Larry William's R%, A/D e CCI. Para a classificação, foi utilizado RNA e SVM para treinamento e classificação dos resultados futuros. Os resultados da classificação por RNA ficaram entre 71,17% e 82,85% de acerto, enquanto os resultados com SVM foram um pouco inferiores, 55,29% e 80,16%, respectivamente, tendo uma grande variação devido aos parâmetros utilizados nos testes e a função utilizada, sendo que a polinomial obteve melhores resultados que RBF (Kara, Acar Boyacioglu, e Baykan 2011).

Patel et al. (2015) publicaram algumas alterações do trabalho de Kara et al. (2011). Com o mesmo objetivo de identificar movimentos futuros de alta ou de baixa, eles estudaram a performance de técnicas diferentes de mineração de dados: Artificial Neural Networks (RNA), *Support Vector Machine* (SVM), *Random Forest* e *Naive-Bayes*. Patel et al. (2015) utilizaram os mesmos parâmetros de entrada utilizados por Kara et al. (2011), ou seja, os indicadores técnicos foram: SMA de 10 dias, WMA de 10 dias, *Momentum*, estocástico K%, estocástico D%, RSI, MACD, Larry William's R%, A/D e CCI.

Transformando os valores diários das ações em números resultantes de indicadores de análise técnica, os autores aplicaram as técnicas de mineração de dados sobre estes valores de duas formas. Na primeira foram utilizados os valores contínuos de cada conjunto de valores resultantes de cada análise técnica, enquanto na segunda forma os autores criaram uma camada que utilizando os mesmos valores de análise técnica obtiveram valores discretos sobre a tendência de cada ação. Chamaram esta camada de “*Trend Deterministic Data Preparation Layer*”. Nesta camada, os dados são coletados e calculados em 10 indicadores técnicos; se o indicador estiver indicando tendência de baixa, o valor da tendência é atribuído -1, se o indicador apontar para uma alta nos preços, é atribuído +1 para o valor da tendência.

Os resultados apontam para uma melhora em relação ao trabalho de Kara, Acar Boyacioglu, e Baykan (2011) na classificação utilizando os valores contínuos e uma melhora ainda maior quando comparado com a técnica dos atributos discretos para o cálculo de tendência. O resultado para o teste com valores contínuos obteve uma taxa de acerto de 78,71% utilizando SVM, enquanto utilizando valores discretos obtiveram um resultado de 89,33%. Os resultados para os valores discretos foram superiores e apontam para uma melhora para todos os quatro métodos de classificação. Além disso, os experimentos mostraram que o método *Random Forest* obteve o melhor resultado com valores contínuos, 83,59%, enquanto *Naive-Bayes* com os valores discretos 90,19% (Patel et al. 2015).

Chang, Fan, e Lin (2011) também estudaram a classificação de movimentos futuros em ações através da geração de regras difusas em árvores (do inglês *Fuzzy Decision Trees* - FDT) para grupos de ações semelhantes. Seu método baseou-se em agrupar as ações em *clusters*, com padrões de movimentos semelhantes. Para isso, foram geradas regras que melhor representam cada grupo. Para melhor assertividade, foi adicionado um processo de otimização dos pesos dos atributos, para a entrada de informações do algoritmo de agrupamento, com um algoritmo genético. Neste processo também foram selecionados os atributos mais importantes na previsão de cada ação através de um algoritmo de análise de regressão passo-a-passo, resultando assim apenas os indicadores técnicos que geram menor erro na previsão, os resultados são apresentados na tabela 3.2. Ao final, regras para cada grupo são avaliadas utilizando um AG e um modelo de classificação por árvore de decisão difusa, para encontrar a regra mais performática para o grupo. Com este método foi possível encontrar regras para negociação de cada grupo de ação e utilizar em futuras decisões. Para a entrada de dados, utilizou-se dos indicadores técnicos e as diferenças entre indicadores técnicos, transformando em osciladores, conforme a tabela 3.1.

Tabela 3.1: Indicadores técnicos selecionados (Chang, Fan, e Lin 2011)

Dados de Entrada	
Indicadores técnicos	SMA de 5 dias Viés entre o preço e a média de 5 dias SMA de 6 dias SMA de 10 dias Estocástico %K de 9 dias Estocástico %D de 9 dias RSI de 6 dias MACD de 9 dias Williams %R de 12 dias

	SMA de 20 dias RSI de 12 dias Diferença entre estocástico %K e %D
Atributos oriundos da diferença entre preço e indicadores técnicos	Diferença SMA de 5 Diferença SMA de 6 Diferença SMA de 10 Diferença do Viés da SMA de 5 Diferença do Viés da SMA 10 Diferença do RSI de 6 dias Diferença Estocástico %K de 9 dias Diferença estocástico %D de 9 dias Diferença do RSI de 12 dias Diferença de Williams R% de 12 dias

Na tabela 3.2 são apresentados os indicadores técnicos e suas diferenças selecionados através do método proposto pelos autores.

Tabela 3.2: Indicadores técnicos selecionados para cada ação (Chang, Fan, e Lin 2011)

Ações	Tipos de Entrada	Atributos de Entrada
Google	Indicador Técnico	Williams R% de 12 dias
	Diferença dos indicadores técnicos	Diferença do viés de 10 dias Diferença do RSI de 6 dias Diferença do RSI de 12 dias
Yahoo	Indicador Técnico	RSI de 12 dias

	Diferença dos indicadores técnicos	Diferença do W%R de 12 dias Diferença do RSI de 6 dias Diferença do RSI de 12 dias
IBM	Indicador Técnico	William's R% de 12 dias
	Diferença dos indicadores técnicos	Diferença do Viés de 10 dias Diferença do RSI de 6 dias Diferença do W%R de 12 dias
GM	Indicador Técnico	William's R% de 12 dias
	Diferença dos indicadores técnicos	Diferença do Viés de 10 dias Diferença do RSI de 12 dias Diferença do W%R de 12 dias
Microsoft	Indicador Técnico	RSI de 6 dias
	Diferença dos indicadores técnicos	Diferença do Viés de 10 dias Diferença do RSI de 12 dias Diferença do W%R de 12 dias

Wang e Shang (2014) propuseram a utilização de *Least Square Support Vector Machine* (LSSVM) para classificar a direção da variação de preços diários. No processo de preparação dos dados para sua proposta, Wang e Shang utilizaram como classe os valores discretos "0" e "1", "0" para caso o dia subsequente fosse abaixo do dia atual, e "1" para alta. Os dados originais foram escalados entre o valor -1 e 1 para garantir que os valores maiores não sobreponham os valores menores no modelo de classificação. E como entrada de valores para o algoritmo, foram utilizados os indicadores técnicos SMA de 10 dias, WMA de 10 dias, estocástico %K, estocástico %D, RSI, MACD, William's R%, A/D e CCI. Seu teste baseou-se no índice *China Security Index 300* e utilizou como comparação os métodos *Probabilistic Neural Network* (PNN), *Liear Discriminant Analysis* (LDA) e *Quadratic Discriminant Analysis* (QDA). Obtendo os resultados apresentados na tabela 3.3.

Tabela 3.3: Resultados do método LSSVM comparados à PNN, QDA e LDA (Wang e Shang, 2014)

	LSSVM	PNN	QDA	LDA
Assertividade no Treinamento	92,97%	92,89%	86,87%	88,18%
Assertividade no Teste	89,12%	80,97%	87,92%	87,31%

Como resultado, o LSSVM obteve melhores resultados comparado ao PNN, QDA e LDA. Wang e Shang realizaram o teste estatístico não-paramétrico de McNemar, apresentado na tabela 3.4, comparando em duplas para verificar o grau de significância do resultado.

Tabela 3.4: McNamar (p-values) para comparação de performance

	PNN	QDA	LDA
LSSVM	0,679	4,654	10,321
PNN		0,327	2,326

A diferença dos resultados entre os modelos aponta que o LSSVM ficou entre 1% e 5% de graus de significância. Não alcançando significância apenas comparado ao PNN com os dados de treinamento, porém alcançou significância com todos os outros dados.

Zbikowski (2015) utilizou uma modificação do SVM, a *Volume Wighted SVM* (VW-SVM), com o objetivo de gerar uma estratégia de negociação. Como entrada de dados foram utilizados indicadores técnicos selecionados através de *Fisher Score* (F-score). Os indicadores técnicos utilizados foram *Average True Range* (ATR), *Vortex Indicator*, *On-Balance Volume* (OBV), *William's %R*, *RSI*. Ele inclui também em seu modelo os valores do desvio padrão e da taxa de retorno.

Como método de avaliação de seu método proposto, Zbikowski criou um algoritmo de simulação de negociações, tendo como entrada os valores diários e através do resultado do algoritmo de classificação,

realiza a compra ou venda da ação. Com esta simulação ele obteve a taxa de retorno nas negociações simuladas. Os resultados obtidos com as técnicas foram avaliados separadamente e progressivamente, para poder avaliar separadamente o resultado de cada uma.

Na tabela 3.5 são expostos os principais resultados. Pode-se verificar que somente com a combinação das técnicas de VW-SVM e F-score é que se obteve um resultado superior à compra e manutenção (C&M). C&M é a compra no primeiro dia de negociações e a manutenção da ação até o último dia, obtendo assim uma comparação do resultado sem negociação entre o período de início e fim.

Tabela 3.5: Comparativo SVM, VW-SVM, SVM + F-score e C&M

Técnica	Taxa de Retorno
SVM	-15,60%
VW-SVM	51,16%
SVM + F-score	35,79%
VW-SVM + F-score	92,93%
Compra e Manutenção (C&M)	78,05%

3.1 Considerações

O trabalho de classificar a direção da variação do preço da ação é bastante complexo, e em muitos casos os autores chegam a resultados somente um pouco melhores do que seria o caminho aleatório. Pode-se ressaltar também que a fase de pré-processamento é essencial, e a combinação de algumas das técnicas como seleção de atributos e AG apresentadas no tópico de estado da arte pode alcançar melhores resultados. Nota-se que algoritmos do tipo ensemble, como o *Random Forest* ainda são pouco explorados neste contexto. A avaliação de resultados através de simulação de negociação torna-se interessante visto que as operações deste mercado são feitas desta maneira e o simples cálculo de taxa de acerto pode esconder médias de ganhos diferentes entre as classificações corretas e incorretas.

Capítulo 4

Método

Conforme os objetivos definidos para este trabalho, será empregado um conjunto de técnicas para classificar se a ação irá subir ou cair no dia seguinte. O método a ser empregado está dividido em duas etapas, na primeira etapa será aplicado o processo de otimização de janelas de cálculo dos indicadores técnicos detalhado na seção 4.2; na segunda etapa será aplicado o processo de otimização de discretização detalhado na seção 4.3. Além disso, na seção 4.4 é apresentado como os AGs são parametrizados e na seção 4.5 é detalhada a forma de avaliação dos métodos empregados. Após, na seção 4.6 é detalhado como será capturada e formatada a base de dados. Mas antes, na seção 4.1 são apresentadas as características dos dados e suas distribuições para que posteriormente as técnicas sejam analisadas e justificadas.

4.1 Características dos Dados

Com o objetivo de analisar as características dos dados para então escolher os algoritmos que irão compor o conjunto de técnicas selecionadas, foram feitos testes preliminares com dados do índice S&P 500 através da biblioteca SciKitLearn e exibidos graficamente através da biblioteca Matplotlib, ambas utilizando Python. As características escolhidas para análise preliminar foram ADX, Williams %R e Estocástico %K.

Observando os histogramas da figura 4.1 é possível identificar que os conjuntos não são paramétricos pois não contêm distribuições normais pois a média, mediana e a moda não são próximas o suficiente para tal afirmação. Além disso, os atributos estão em escalas diferentes e neste caso, para aplicar este conjunto

de dados em aprendizados de máquina baseados em distância é necessário que o conjunto seja normalizado ou tenha seus valores padronizados.

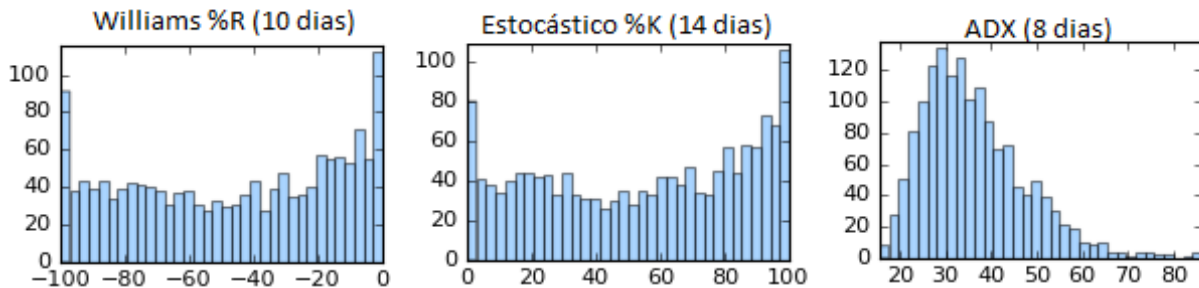


Figura 4.1: Histograma dos atributos Williams %R, Estocástico %K e ADX

Visualização de Superfícies de Decisão

Dado que as características do conjunto são características com escalas diferentes, os testes a seguir têm a necessidade de padronizar as escalas dos atributos para a utilização de algoritmos baseados em distância. Com o intuito de analisar os algoritmos mais adequados para o tipo de informação foram realizados testes para a visualização das superfícies de separação geradas pelos algoritmos de aprendizado de máquina utilizando uma amostra de 100 instâncias do mesmo conjunto de dados do índice S&P 500, com as mesmas três características supracitadas. Essas três características foram separadas em dois testes, o primeiro teste foi o cruzamento entre ADX e Estocástico %K e segundo foi o cruzamento entre ADX e Williams %R.

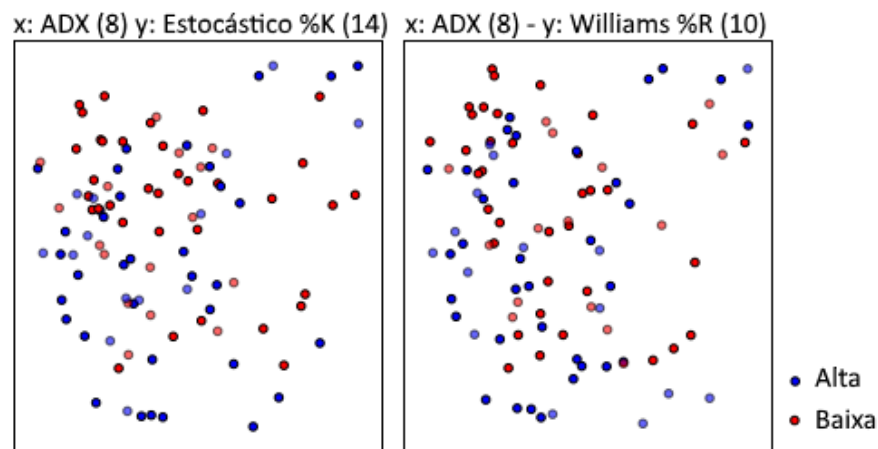


Figura 4.2: Cruzamento dos atributos ADX com Estocástico %K e ADX com Williams %R

Como pode-se observar na figura 4.2 os dados podem não obedecer uma separação linear, já que sua característica ruidosa e caótica impede uma escolha de um classificador linear. Para melhor esclarecimento de qual técnica pode adaptar-se ao tipo de informação, ambos os cruzamentos apresentados na figura 4.2 foram aplicados aos seguintes algoritmos de aprendizado de máquina: *CART Decision Tree*, *Random Forest*, *Gaussian Naive-Bayes*, *k-Nearest Neighbors*, *Linear SVM*, *RF SVM* e *Polynomial SVM* e são apresentados nas figuras 4.3 e 4.4.

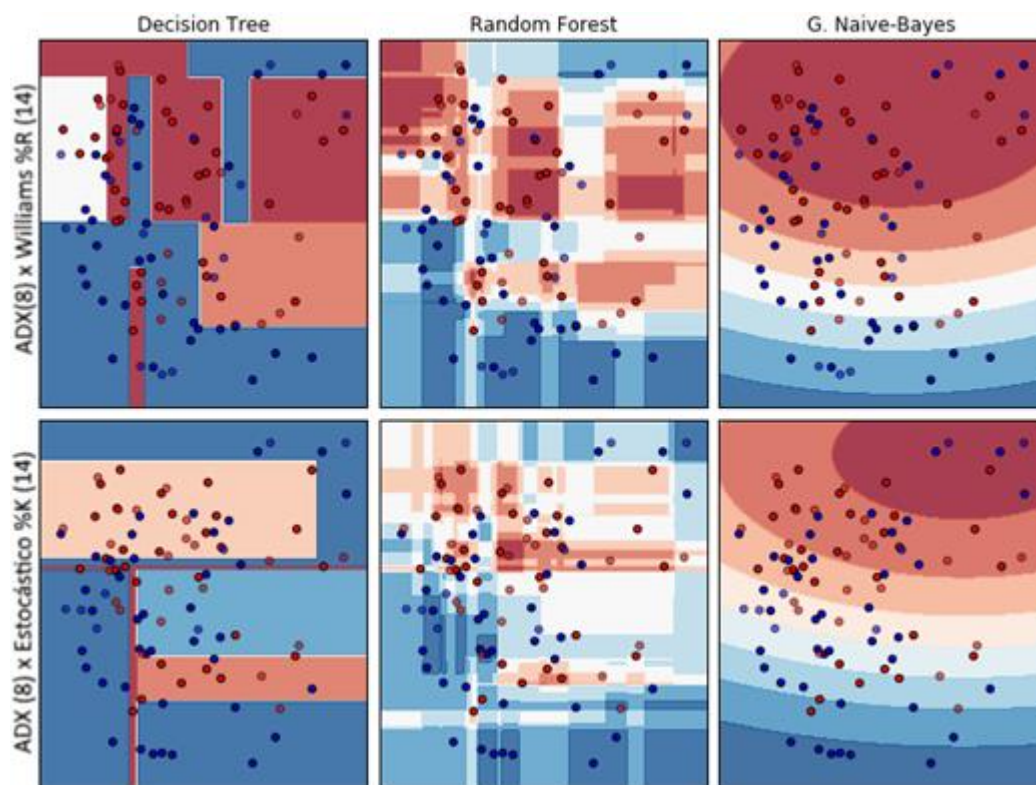


Figura 4.3: Superfícies de decisão dos algoritmos *CART Decision Tree*, *Random Forest*, *Gaussian Naive-Bayes*

Os algoritmos apresentados nas figuras 4.3 e 4.4 evidenciam que estruturas de decisão como CART tem melhor adaptação para tal conjunto de dados, assim como algoritmos baseados em densidade, como o kNN, e radiais como o algoritmo SVM RBF. Por outro lado, algoritmos que se adaptam bem à dados paramétricos, como é o caso do *Gaussian Naive-Bayes*, não obtiveram bons resultados. Outro algoritmo que

também não apresentou uma boa superfície de decisão para o conjunto de dados foi o *Linear SVM*, pois este algoritmo despreza ilhas classes geradas pela interação dos atributos em faixas.

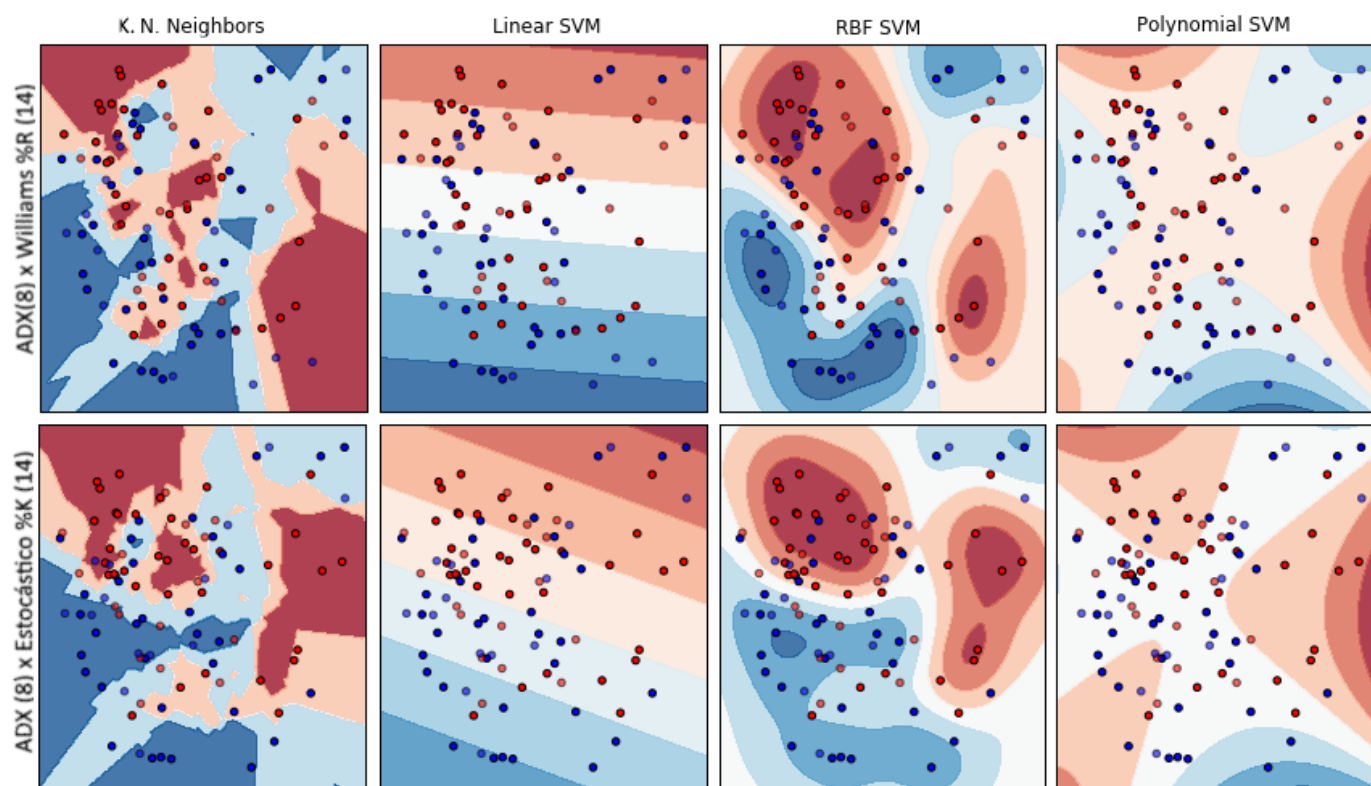


Figura 4.4: Superfícies de decisão dos algoritmos *k-Nearest Neighbors*, *Linear SVM*, *RBF SVM* e *Polynomial SVM*

Assim, com tais características demonstradas do conjunto de dados, é necessário que a aplicação de algoritmos de classificação seja capaz de separar as informações em várias “ilhas” de decisão, se adaptando às características não paramétricas, à interação entre atributos e aos dados excessivamente ruidosos que podem gerar armadilhas de generalização com o conjunto de treinamento causando alta variância dos resultados. SVM e *Random Forest* são algoritmos que se comportam bem com tais condições, sendo que o *Random Forest* por sua característica de ser um algoritmo do tipo *ensemble* e por sua natureza aleatória de atributos e instâncias, consegue evitar armadilhas de generalização e alta variância dos resultados. Por outro lado o SVM tem em sua essência a minimização dos riscos estruturais, maximizando através de sua margem de superfície de decisão a generalização do modelo treinado. Contudo, este trabalho realizará seus

experimentos com *Random Forest*, que como já experimentado por Patel et al. (2015), é um algoritmo que se sobressaiu em relação ao SVM em seus experimentos.

Além da escolha do algoritmo de classificação, este trabalho propõe a utilização de um algoritmo de seleção de atributos para reduzir a dimensionalidade do conjunto de dados original. Dentre os algoritmos de seleção de atributos, o algoritmo ReliefF se adapta a todos os requisitos supracitados na análise das características dos dados. Sendo este um algoritmo robusto para a interação entre atributos, e também tolerante a ruídos e a dados não-paramétricos.

4.2 Algoritmo para Otimização de Janelas de Tempo

A maioria dos indicadores técnicos apresentados no capítulo 2 recebem por parâmetro a janela de tempo representado pela variável n para efetuar o seu cálculo. Esta janela de tempo é quantidade de dias no passado que este indicador irá analisar para encontrar o valor do indicador no momento t , conforme exemplificado na figura 4.5 na qual é apresentado o calcula de uma média móvel simples no período t .

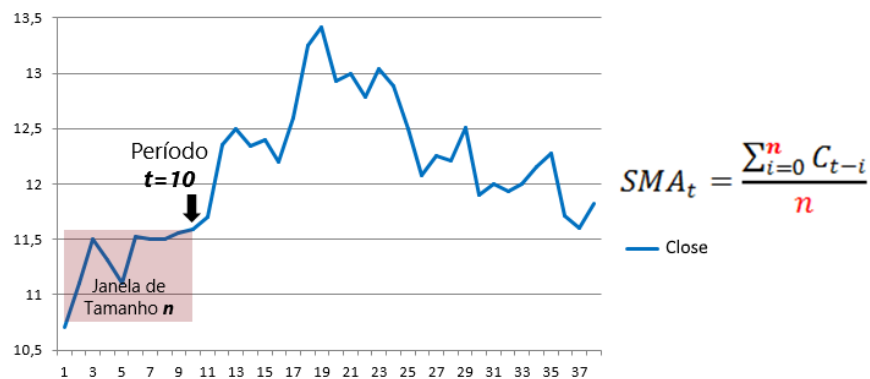


Figura 4.5: Exemplo de utilização da variável n na fórmula SMA

Analistas de mercado entendem que quanto maior a janela de tempo que se analisa, maior é o prazo para investimento. Por se tratar de um algoritmo que decide dia-a-dia a compra e venda, entende-se por uma negociação de curto prazo e em alguns casos o n indicado pelo autor do indicador pode não representar a melhor performance para o classificador proposto neste trabalho. Para encontrar o valor ótimo global, ou

próximo do ótimo global, para cada variável n aplicado ao problema do presente trabalho, foi proposto a aplicação de um algoritmo meta-heurístico para otimização de janelas de tempo para o cálculo de indicadores técnicos. Para a realização desta tarefa, é necessário um algoritmo meta-heurístico para otimização combinatória mono-objetivo que realize a busca em espaços complexos, não-sequencial e com relevante grau de aleatoriedade em sua busca, visto que os dados não seguem uma progressão linear ou gradiente. Para executar tal tarefa de busca com otimização a nível global, algoritmos genéticos são adequados, assim como *Particle Swarm Optimization* (PSO) e suas variações. Em algumas pesquisas os Algoritmos Evolutivos como AGs têm obtido desempenhos melhor do que outras estratégias de busca (De Toro et al. 2006). Algumas características destes métodos merecem ser destacadas: flexibilidade, generalidade, exploração, exploração, paralelismo, capacidade de escapar de ótimos locais, são menos suscetíveis à forma ou continuidade das funções (Bäck, Fogel, e Michalewicz 2000a; Bäck, Fogel, e Michalewicz 2000b). Por outro lado, o PSO é um importante algoritmo com boas qualidades para diversos problemas, contudo, a busca pelo melhor conjunto de janelas de tempo contém um espaço de busca com variações extremas de resultado sem uma convergência gradiente de tamanho de janelas tempo em alguns pontos. Dado esta questão e tendo em vista que um dos problemas do PSO é a provável convergência prematura das partículas à um ponto, pois a aleatoriedade de busca ocorre apenas no início, após isso as partículas tendem a se agrupar, o algoritmo meta-heurístico escolhido foi o AG, cuja a sua otimização pode ser expressa pela equação 4.1, sendo y o valor ótimo a ser alcançado através da maximização da função *fitness* tendo como parâmetro a variável x que representa o cromossomo.

$$y = \max \{fitness(x)\} \quad (4.1)$$

Então, o algoritmo proposto nomeado de Algoritmo Genético para Otimização de Janelas de Tempo (AGOJT) buscará a combinação do melhor conjunto de janelas de tempo, para isso colocará em cada gene do cromossomo candidato um valor inteiro à ser otimizado pelo processo proposto. Para avaliar se o tamanho de cada janela é o mais adequado é necessário efetivamente submeter o conjunto de dados gerado com os tamanhos de janelas propostos ao classificador e avalia-los. Para alcançar esta etapa o AG proposto e apresentado na figura 4.6 embutirá em seu processo para calcular o melhor conjunto de janelas as seguintes etapas:

- A transformação dos atributos em atributos de indicadores técnicos de acordo com a janela de tempo sugerida em cada gene do cromossomo;
- Seleção de atributos através do algoritmo ReliefF, proposto por (Kononenko 1994), tem a função de reduzir a dimensionalidade dos conjuntos de dados;
- Classificador *Random Forest*, proposto por (Leo Breiman 1994), irá classificar as instâncias de teste;
- E finalmente, a *Fitness Function* será o retorno da taxa de acerto da classificação de um algoritmo *Random Forest*.

Nesta etapa, os dados processados serão apenas de origem contínua.

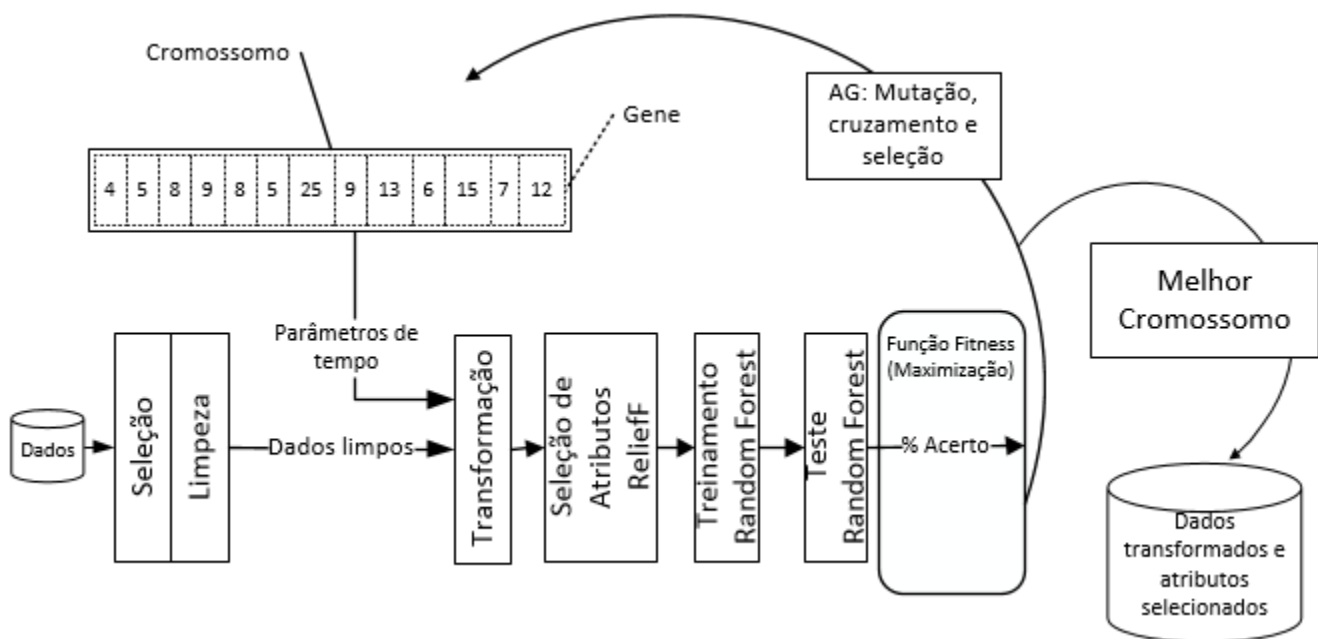


Figura 4.6: Primeira etapa - AG para otimização de janelas de tempo

Seleção e Limpeza

Os dados são selecionados de 01/01/2000 até 31/12/2015. As etapas de limpezas necessárias são:

- Remover valores duplicados.

- Ajustar os valores de mínimo, máximo e abertura quando houver distribuição de dividendos, para isso é necessário utilizar um atributo dos dados originais, o atributo *Adjusted Close*. Este atributo calcula um novo valor do fechamento (atributo *Close*) para que não haja sobressaltos gerados eventos como distribuição de dividendos, *split* e outras situações que alteram o valor da ação sem a negociação do mercado. Em um Split, por exemplo, pode reduzir pela metade o valor de uma ação sem que de fato tenha caído, apenas os investidores terão o dobro de ações enquanto estas ações custarão a metade do preço. A geração de dividendos também reduz o preço do valor que foi distribuído. Por isso, o ideal é utilizar o *Adjusted close* (fechamento ajustado) como valor de referência para o fechamento, porém, há alguns indicadores de análise técnica que utilizam outros atributos, como o valor máximo (*high*), mínimo (*low*) e abertura (*open*); estes atributos não são fornecidos com o valores ajustados e por isso precisam ser ajustados nesta etapa.

Para cada mudança identificada no fator de ajuste, deve-se calcular a taxa de mudança (equação 4.2) e aplicar sobre os atributos *high*, *low* e *open* de todas as instâncias o fator de alteração gerado por estes eventos de mercado (equações 4.3, 4.4 e 4.5).

$$fator = 1 - \left(\frac{close - adjClose}{close} \right) \quad (4.2)$$

$$low = low * fator \quad (4.3)$$

$$high = high * fator \quad (4.4)$$

$$open = open * fator \quad (4.5)$$

Assim como os valores de fechamento (*close*) devem receber os valores fechamento ajustado (*adjusted close*). O resultado esperado para esta operação é que não haja discrepâncias não naturais de valores entre os valores de abertura, máximo, mínimo e fechamento ao longo dos anos.

- Outro ajuste necessário no tratamento dos dados é a ordenação dos valores em ordem cronológica. Em muitas fontes de informação os dados são fornecidos da mesma maneira que são apresentados em aplicações voltadas para este mercado, os valores mais recentes primeiro. Porém para o cálculo dos indicadores técnicos os valores originais devem estar em ordem cronológica.

Transformação

Esta é a etapa do processo a ser otimizada através de um algoritmo genético. O AG por meio de seus processos genéticos como mutação e cruzamento, irá informar à função de transformação quais tamanhos de janelas serão calculados para cada atributo, então esta função efetuará o cálculo de cada atributo para todas as instâncias, incluindo as instâncias de teste e avaliação.

Porém, nem todos os indicadores necessitam de uma janela de análise. No caso do presente trabalho o indicador A/D utiliza apenas informações da última cotação, ou seja, não há janela de tempo. Além disso, os grupos de indicadores (ADX, +DI, -DI) e (AroonUp, AroonDown) são indicadores calculados com a mesma janela de tempo. Os indicadores técnicos que terão suas janelas otimizadas pelo AG, e por consequência se tornarão genes dos cromossomos, são os apresentados na tabela 4.1. Ao todo são 13 genes, sendo que são necessários três genes para o cálculo do indicador MACD, pois utiliza-se de três janelas de tempo para o seu cálculo. A tabela 4.1 também indica os valores mínimos e máximos que o AG poderá otimizar, evitando assim valores aleatórios fora do normal para este processo. Em sua maioria, as faixas de janelas foram estendidas para baixo em relação aos valores padrões, pois em quase todos os casos há uma tendência de obtenção de resultados mais relevantes quando utilizados tamanhos de janelas menores do que o padrão. Maiores detalhes sobre a codificação do processo de AG estão descritos na seção 4.4.

Tabela 4.1: Indicadores técnicos que terão a janela de tempo otimizados

	Valor Mínimo - Valor Máximo	Valor Padrão
William's %R	7 – 15	14
<i>Stochastic %K</i>	7 – 15	14
<i>Stochastic %D</i>	3 – 6	3
MACD <i>fast</i>	3 – 9	9

MACD <i>slow</i>	20 – 28	26
MACD <i>Oscillator</i>	8 – 15	12
CCI	6 – 20	20
RSI	6 – 16	14
PSY	10 – 16	14
Aroon	20 – 28	25
ADX (usado em ADX, +DI e -DI)	6 – 15	14
SMA	7 – 13	10
WMA	7 – 13	10

Ao final desta etapa, serão calculados quinze indicadores técnicos que servirão de entrada no processo de seleção de atributos, para sejam filtrados apenas os mais relevantes.

Seleção de Atributos

Para a execução do processo de seleção de atributos, será selecionado uma quantidade fixa de atributos. Alguns trabalhos utilizam todos os atributos alterando apenas o peso na entrada do classificador. Porém, este trabalho pretende selecionar os atributos com melhor ganho de informação. A quantidade foi definida através de testes preliminares, mostrados na figura 4.7.

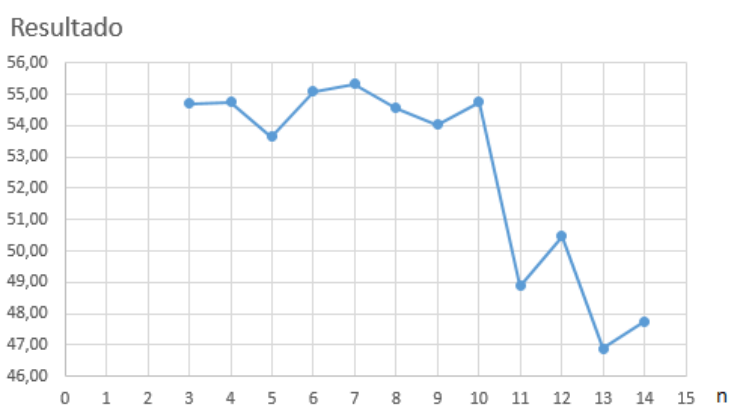


Figura 4.7: Gráfico de resultados rodados com *Random Forest* 80 árvores variando apenas o parâmetro n do Algoritmo ReliefF

A figura 4.7 apresenta o gráfico com os resultados preliminares, utilizando *Random Forest* com tamanhos de janelas prefixadas, o gráfico visa apenas demonstrar o comportamento conforme o número de atributos (variável n do algoritmo ReliefF). Pode-se notar que, após o valor de $n = 0$ os resultados tornam-se absolutamente aleatórios. Dado o gráfico, este trabalho utilizou apenas 7 atributos dos 15 iniciais.

Treinamento

E por último, a configuração do algoritmo *Random Forest* será de 80 árvores ($n = 80$), que conforme experimentos feitos por Patel et al (2015), *Random Forest* com 80 árvores obteve bons resultados comparados com SVM, *Naive-Bayes* e RNA. Para a configuração de número de atributos de cada árvore foi estabelecido o padrão aleatório, ou seja, o número não é fixo variando aleatoriamente entre 1 e 7 para cada uma das 80 árvores.

4.3 Algoritmo para Otimização de Atributos Discretos

Diferentemente de Kim 2000 que propôs um AG para discretização que procurava os melhores limites para intervalos das categorias de cada atributo e Debie, Shafi, Merrick, & Lokan (2014) que utilizaram AG para otimizar a discretização em uma abordagem de deslocamento de janela incremental para encontrar os melhores limiares, esta segunda etapa deste trabalho consiste em utilizar o melhor conjunto de atributos encontrados na primeira etapa e encontrar o melhor formato para cada um desses atributos. Para isso, foi escolhido AG novamente, motivado pela utilização em trabalhos pregressos, nomeado de Algoritmo para Otimização de Atributos Discretos (AGOAD). Então esta etapa irá otimizar o formato de cada atributo nos seguintes valores:

- Discreto utilizando EWD;
- Discreto utilizando EFD;
- Contínuo.

O AG ainda irá selecionar quantas divisões serão obtidas para cada atributo discreto. Os valores estarão entre 3 e 12. Esta é uma otimização com genes compostos entre o formato do atributo e suas divisões, conforme pode ser visualizado na figura 4.6. Porém com a função objetivo sendo a mesma do AGOJT, conforme apresentado na equação 4.6, sendo y a solução ótima e x o cromossomo.

$$y = \max \{fitness(x)\} \quad (4.6)$$

O processo incluindo as duas etapas pode ser melhor compreendido através da figura 4.7, onde demonstra que o ponto de interconexão entre os dois AG é a base de atributos transformados com o melhor resultado na primeira etapa.

Ao final desta etapa será possível avaliar o resultado, comparando se o processo de otimização de atributos discretos pode melhorar a taxa de classificação e também será possível executar um algoritmo de negociação a fim de simular em uma situação mais próxima da negociação dos mercados, conforme apresentado no seção 4.5.

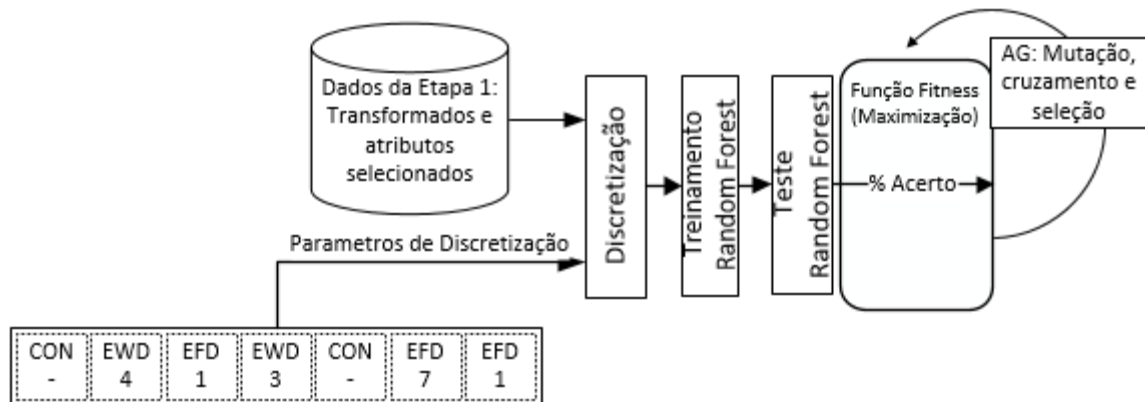


Figura 4.8: Segunda etapa, otimização de discretização

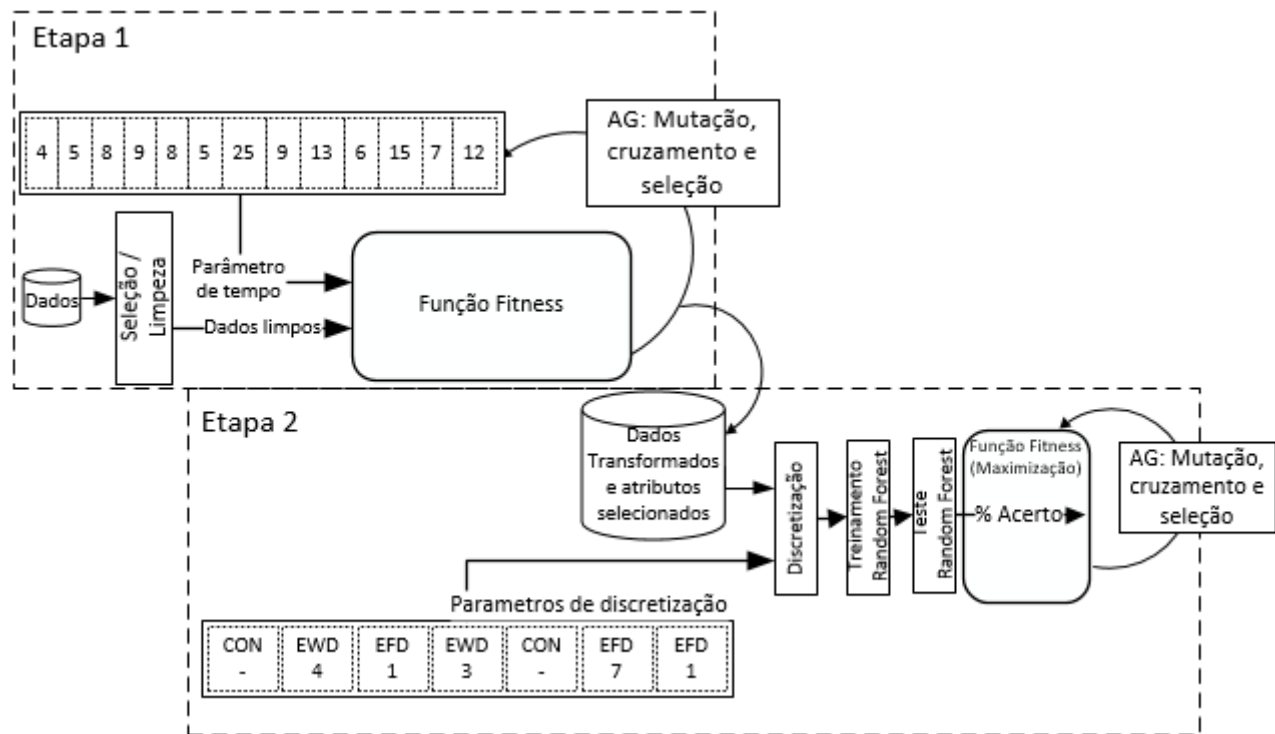


Figura 4.9: Etapas 1 e 2 integradas

4.4 Configura  o dos AGs

Os algoritmos gen ticos das etapas 1 e 2 s o do tipo *Elitism Selection*, no qual seleciona os mais bem-sucedidos para a pr xima gera  o, evitando que indiv duos com baixo resultado na fun  o *Fitness* sejam mantidos para a gera  o seguinte. A configura  o deste algoritmo consiste nos seguintes par metros solicitados em um AG Elitista:

- Popula  o: 25;
- Taxa de muta  o: 35% das caracter sticas;
- Operador de muta  o: 1/12 dos indiv duos;
- Taxa de elitiza  o: 90% da popula  o, 10% dos inferiores s o removidos;
- M ximo de evolu  es (crit rio de parada): 500.

4.5 Avaliação dos Resultados

Conforme definido nos objetivos do presente trabalho, a avaliação se dará pela taxa de acerto e pelo resultado do algoritmo de negociação.

4.5.1 Taxa de Acerto

A taxa de acerto ou precisão é uma medida que é apurada como o total de classificações corretas dividido pelo total de instâncias classificadas. A taxa de acerto será utilizada nas duas etapas pela função *Fitness* para encontrar o melhor cromossomo e na fase de validação dos modelos gerados utilizando os 30% finais da base de dados.

4.5.2 Algoritmo de Negociação

A proposta de avaliação baseada no resultado da negociação é uma forma de avaliar o comportamento de todo o processo em uma simulação de negociações da bolsa de valores e uma forma de comparar os resultados com o trabalho de Żbikowski (2015) apresentado no estado da arte do presente trabalho.

O algoritmo baseia-se no modelo de classificação gerado pela fase de treinamento do *Random Forest* para classificar se a ação irá subir ou não no dia seguinte e assim gerar o sinal de compra ou não da ação. Para isso, o melhor modelo gerado pelos AGs das etapas 1 e 2 de cada ação será submetido ao algoritmo de negociação para simular o desempenho. Além disso, para reduzir o número de negociações o algoritmo proposto não irá utilizar a classificação binária; em vez disso, ele utilizará a distribuição de probabilidade superior a 60% para a classe, calculada para cada exemplo, com base no modelo de formação apresentado na Figura 4.5. Os classificadores baseados em árvores de decisão, como o *Random Forest*, utilizam distribuição de probabilidade para a definição de qual classe será atribuída para cada instância. Este trabalho baseia-se no problema de duas classes, o que significa que, a classe com probabilidade acima de 50% é a escolhida. Para a *Random Forest*, a distribuição de probabilidade é calculada pela média de todas as árvores construídas pelo algoritmo.

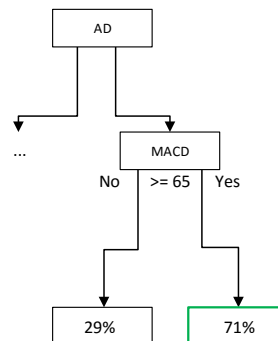


Figura 4.10 - Probabilidade de distribuição em árvores de decisão

O algoritmo proposto realizará a negociação apenas quando a probabilidade do classificador for superior a 60%, reduzindo o número de negociações. O algoritmo será executado sobre o conjunto de dados de validação, ou seja, os últimos 30% dos dados. Para cada instância, o algoritmo irá escolher se compra, vende, mantém-se comprado ou mantém-se sem comprar:

- Compra a ação quando a classificação indicar que a ação irá subir no dia seguinte e tiver dinheiro em caixa, comprando a quantidade máxima de ações que a quantidade de dinheiro permitir;
- Vende todas as ações quando a classificação apontar que o dia seguinte a ação irá cair e se tiver ações compradas;
- Mantém as ações compradas se a classificação apontar para uma alta no dia seguinte e o algoritmo tiver as ações compradas;
- Mantém sem comprar ações se a classificação apontar para queda no próximo dia e não tiver ações compradas.

O algoritmo começa com US\$100.000,00 em caixa, disponíveis para comprar ações. Para efetuar a comparação de forma consistente, o resultado acumulado do algoritmo deve ser reduzido do resultado do C&M, sendo assim, a comparação leva em conta apenas o que o algoritmo fez a mais do que teria feito se não tivesse negociado a ação, apenas mantida durante todo o período.

4.6 Base de Dados

Os dados utilizados para o processamento do algoritmo proposto são 15 anos de duas ações da bolsa de valores Dow Jones: Exxon Mobil (XOM) e Microsoft (MSFT) e um índice acionário: S&P 500 (SP500), sendo assim, serão três conjuntos de dados para aplicar o método proposto. O período selecionado inicia-se em janeiro de 2000 e finaliza em dezembro de 2014. Cada conjunto de dados será dividido em:

- 40% para treinamento do classificador;
- 30% utilizado na fase de testes dentro do algoritmo AG;
- E outros 30% para ser utilizado na fase de validação do modelo gerado e para o algoritmo de negociação.

Os períodos das divisões do conjunto de dados são apresentados na figura 4.1



Figura 4.11: Divisão da base de dados

Os dados para estas bases podem ser obtidos através da página de notícias financeiras do Yahoo, o *Yahoo Finance*, sendo disponibilizados através de arquivos no formato CSV.

4.7 Considerações

O modelo apresentado explora duas otimizações importantes para a predição de valores. A primeira otimização explora o tamanho das janelas dos indicadores técnicos, cujos valores padrões podem ser compatíveis com o ritmo de operações que serão realizadas pela proposta. Com este algoritmo genético proposto, ainda não proposto por nenhum trabalho da área, isto será contornado. Outro ponto importante, a avaliação por negociação, parecida com a avaliação proposta por Zbikowski, porém este modelo leva em consideração o percentual da probabilidade de distribuição. Com esta avaliação poderá se ter maior compreensão do que o modelo gerado pode alcançar em termos de lucro, fazendo uma simulação mais próxima da realidade deste mercado.

Capítulo 5

Resultados Obtidos

Para desenvolvimento do método proposto, foi utilizado a linguagem e plataforma Java SE 1.7 juntamente com as seguintes ferramentas:

- Biblioteca do Weka 3.7.10, para executar as tarefas de classificação e seleção de atributos;
- Biblioteca TA-Lib, para realizar o cálculo dos atributos de indicadores técnicos, provendo a maioria dos indicadores necessários;
- *Framework* JGAP, para implementação dos algoritmos genéticos das etapas 1 e 2.

5.1 Algoritmo Genético para Otimização de Janelas de Tempo

Os resultados foram obtidos nas duas etapas, para que se possa avaliar a influência de cada otimização. Os resultados da seleção de atributos e dos tamanhos das janelas da primeira etapa são apresentados nas tabelas 5.1, 5.2 e 5.3. Nestas tabelas estão os sete atributos selecionados pelo algoritmo ReliefF em ordem de seleção, o tamanho da janela obtida no processo e os valores padrões das janelas para que se possa comparar.

Tabela 5.1: Atributos selecionados e tamanhos de janelas para Microsoft.

	1°	2°	3°	4°	5°	6°	7°
Atributos	ADX	Stoch%K	A/D	Williams %R	Psy	RSI	MACD
Tamanho da Janela	6	11	-	10	12	7	(6,8,22)
Tamanho Padrão da Janela	14	14	-	14	14	14	(9,12,26)

O indicador técnico Accumulation/Distribution (A/D) não usa uma janela de tempo em sua formulação. Por outro lado, o MACD usa três janelas de tempo para o seu cálculo, como apresentado na tabela 5.1.

Tabela 5.2: Atributos selecionados e tamanhos de janelas para Exxon Mobil

	1°	2°	3°	4°	5°	6°	7°
Atributos	ADX	CCI	Stoch%K	Williams %R	Aroon Up	-DI	RSI
Tamanho da Janela	7	9	14	11	25	7	7
Tamanho Padrão da Janela	14	10	14	14	25	14	14

Tabela 5.3: Atributos selecionados e tamanhos de janelas para S&P 500

	1°	2°	3°	4°	5°	6°	7°
Atributos	ADX	Aroon Up	RSI	Williams %R	WMA	-DI	Stoch %K
Tamanho da Janela	6	25	7	10	6	6	15
Tamanho Padrão da Janela	14	25	14	25	10	14	14

A fase de testes do modelo do AGOJT calcula a taxa de acerto (precisão) para a previsão do movimento no dia seguinte, como demonstrado na tabela 5.4. Além disso, cada modelo construído pelo AGOJT foi testado sobre a base de validação, ou seja, o mesmo conjunto de dados que será utilizado pelo algoritmo de negociação. Além dos testes com o conjunto de validação, também foi calculado com o tamanho de janela padrão, para que se possa comparar a otimização feita pelo AGOJT.

Tabela 5.4: Resultados AGOJT e comparações

Ação	Precisão d e AGOJT	Precisão AGOJT com Conjunto de Dados de Validação	Precisão com Tamanho de Janela Padrão	Precisão com Tamanho de Janela Padrão e com Conjunto de Dados de Validação
SP500	60.77%	57.68%	54.59%	50.44%
MSFT	58.74%	52.38%	54.06%	50.00%
XOM	59.01%	53.00%	50.53%	49.29%
Média	59.51%	54.35%	53.06%	49.21%

O modelo resultante do AGOJT foi utilizado para a simulação de negociações proposta no tópico 4.5.2 para classificar o movimento do próximo dia. Aplicado para cada conjunto de dados de validação. Os resultados da negociação são apresentados na tabela 5.5 comparados com C&M.

Tabela 5.5: Resultados do algoritmo de negociação de AGOJT comparado C&M

Ação	Acumulado C&M (US\$)	Acumulado AGOJT (US\$)	Lucro de AGOJT sobre C&M (US\$)	Lucro de AGOJT sobre C&M (%)
SP500	200.270,39	374.871,92	+ 174.601,53	+87,18%
MSFT	220.390,33	320.029,56	+ 99.639,23	+45,21%
XOM	180.824,16	297.202,20	+ 116.378,04	+64,36%
Média	200.494,96	333.701,23	+ 130.206,27	+66,58%

Com o intuito de comparar os ganhos utilizando o AGOJT, cada conjunto de dados foi executado com o tamanho de janelas padrão de cada indicador técnico - valores padrões de janelas foram apresentados

nas tabelas 5.1, 5.2 e 5.3. O desempenho do algoritmo sem AGOJT e com valores contínuos são apresentados na tabela 5.6.

Tabela 5.6: Tamanhos de janela padrão comparados com C&M

Ação	Acumulado - Tamanho de janela padrão (US\$)	Lucro sobre C&M (US\$)	Lucro sobre C&M (%)
SP500	153.718,89	-46.551,50	-23,24%
MSFT	231.781,88	11.391,56	+5,17%
XOM	128.737,96	-52.086,20	-28,80%
Média	132.746,23	-29.086,04	-15,62%

Como apontado na tabela 5.5, o algoritmo AGOJT superaram C&M em 66.58% na média. Além disso, comparando com a tabela 5.6, a otimização feita pelo AGOJT é de 82% superior comparado com o tamanho de janelas padrão em média. Visto os resultados desta etapa, poderemos comparar estes resultados no próximo tópico com o algoritmo completo.

5.2 Resultados do Processo Completo

O processo completo apresentado na figura 4.3 é a combinação dos dados transformados com as melhores janelas para os indicadores técnicos encontrados no AGOJT com algoritmo AGOAD. O algoritmo AGOJT encontra o melhor conjunto de atributos, extraídos dos dados não processados, deixando sete atributos dos quinze inicialmente processados. Na segunda etapa, a etapa do AGOAD, o algoritmo proposto constrói e avalia modelos através da discretização dos atributos gerados na etapa anterior, definindo o melhor formato para o atributo entre discretizado com EWD, EFD ou contínuo. Além disso, o AGOAD também otimiza o número de divisões que o atributo categórico terá. Os resultados de precisão são apresentados na tabela 5.7. Já na tabela 5.8 são apresentados os resultados da negociação utilizando o modelo completo. E por último, os formatos de atributos encontrados no AGOAD são apresentados nas tabelas 5.9, 5.10 e 5.11.

Tabela 5.7: Precisão no processo completo, AGOJT e AGOAD

Ação	Precisão - Base de Teste	Precisão - Base de Validação
SP500	60.86%	57.59%
MSFT	59.01%	55.30%
XOM	58.21%	57.59%
Média	59.36%	56.83%

Tabela 5.8: Resultados de negociação no processo completo, AGOJT e AGOAD

Ação	Acumulado Negociação	Lucro sobre C&M (US\$)	Lucro sobre C&M (%)
SP500	US\$ 409,684.59	+ US\$ 209,414.21	104,57%
MSFT	US\$ 355,367.95	+ US\$ 134,977.62	61,24%
XOM	US\$ 263,799.29	+ US\$ 82,975.13	45,59%
Média	US\$ 342,950.61	+ US\$ 142,455.65	70,57%

Na média, os resultados apresentados pelo processo completo são superiores ao da primeira etapa. A comparação expõe que apenas o resultado da Exxon Mobil não aumentou. Também é possível observar nas tabelas 5.9, 5.10 e 5.11 que o algoritmo AGOAD selecionou mais atributos discretos que contínuos, ao todo são 8 EFD, 8 EWD e 5 contínuos.

Tabela 5.9: Formato dos atributos encontrados no AGOAD para S&P 500

Atributo	Formato	Número de Categorias
1º	Contínuo	-
2º	Contínuo	-
3º	Contínuo	-
4º	EFD	9
5º	EWD	5
6º	EFD	7
7º	EWD	8

Tabela 5.10: Formato dos atributos encontrados no AGOAD para Microsoft

Atributo	Formato	Número de Categorias
1º	EWD	5
2º	EFD	10
3º	EWD	7
4º	Contínuo	-
5º	EWD	4
6º	EWD	8
7º	EWD	7

Tabela 5.11: Formato dos atributos encontrados no AGOAD para Exxon Mobil

Atributo	Formato	Número de Categorias
1º	EFD	8
2º	EWD	3
3º	EFD	11
4º	EFD	4
5º	EFD	9
6º	Contínuo	-
7º	EFD	4

Para maiores detalhes sobre o algoritmo de negociação, pode-se analisar graficamente o desempenho do algoritmo da primeira etapa comparado com o C&M nas figuras 5.1, 5.3 e 5.5 e os resultados do algoritmo de negociação para o modelo gerado pelo processo completo através das figuras 5.2, 5.4 e 5.6, está representado em preto o desempenho C&M e cinza o desempenho dos algoritmos.

Os resultados obtidos podem ser comparados com os resultados obtidos por Żbikowski (2015), apresentado na tabela 5.12. Żbikowski (2015) executou o algoritmo de negociação em um período de 10 anos, em média o algoritmo de negociação que utilizou o modelo de classificação da direção do dia seguinte obteve lucro de 93,93% e 14,88% acima da C&M. Já para o teste com um modelo de classificador para classificar o movimento cinco dias a frente obteve 168,71% de lucro e 90,66% acima da C&M. No presente

trabalho, utilizando um período de 4,5 anos, em média foi obtido 242,95% de lucro e 70,57% de lucro acima da C&M.

Tabela 5.12: Comparação dos resultados

Autor	Dias a frente	Lucro sobre Investimento	Lucro acima da C&M
Żbikowski (2015)	1	93,93%	14,88%
Żbikowski (2015)	5	168,71%	90,66%
Norberto Ritzmann Júnior	1	242,95%	70,57%

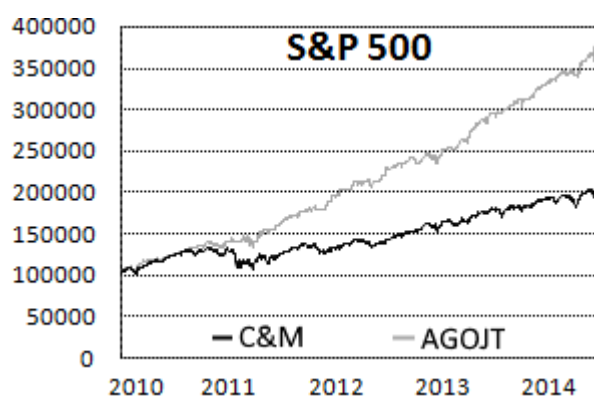


Figura 5.1: Histórico de negociações utilizando AGOJT para o índice S&P 500 comparado com C&M

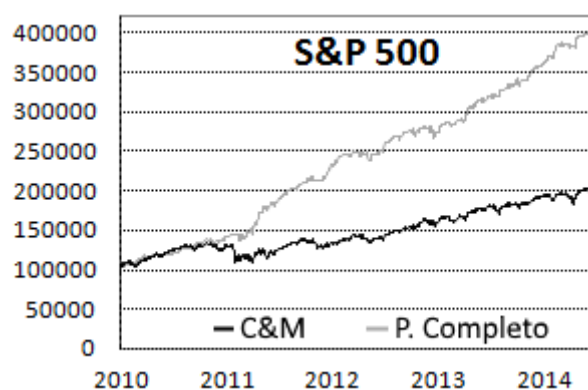


Figura 5.2: Histórico de negociações utilizando o processo completo para o índice S&P 500 comparado com C&M

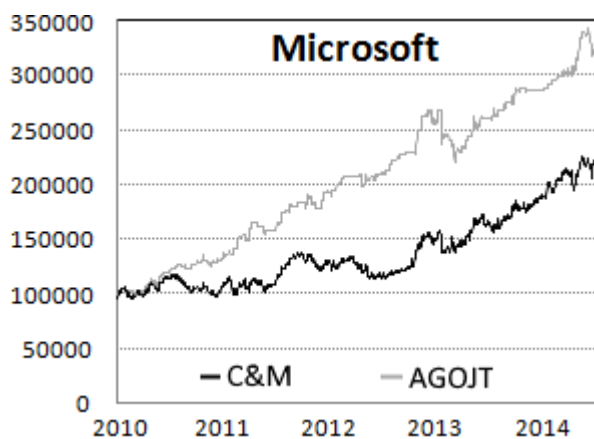


Figura 5.3: Histórico de negociações utilizando AGOJT para a ação Microsoft comparando com C&M

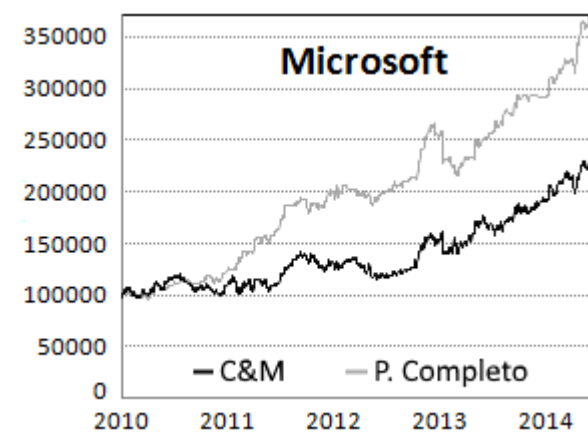


Figura 5.4: Histórico de negociações utilizando o processo completo para a ação Microsoft comparado com C&M

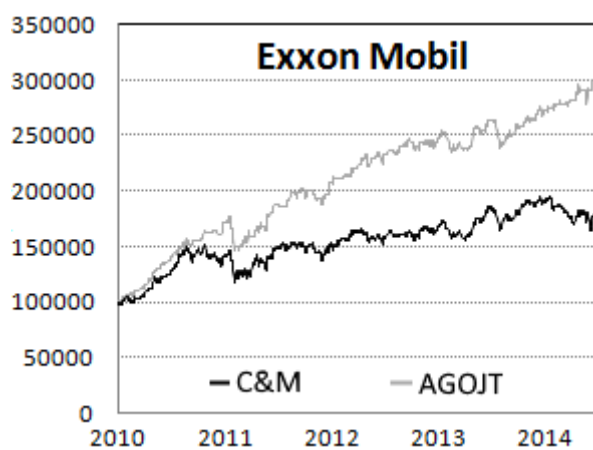


Figura 5.5: Histórico de negociações utilizando AGOJT para a ação Exxon Mobil comparado com C&M

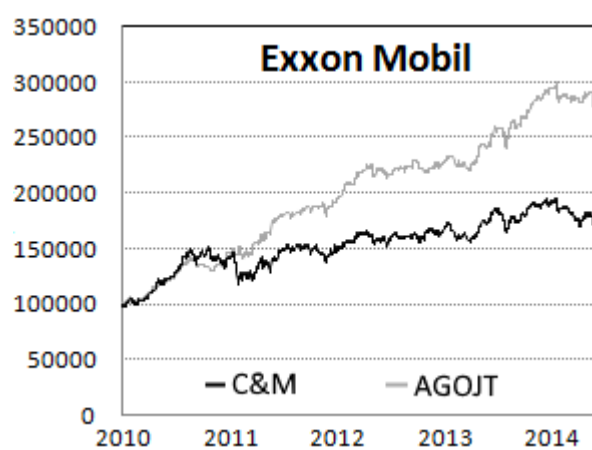


Figura 5.6: Histórico de negociações utilizando o processo completo para a ação Exxon Mobil comparado com C&M

Há de se notar também que os resultados do algoritmo de negociação são influenciados por uma tendência de alta presente em quase todo o período, mesmo assim, o algoritmo supera em todas as situações o C&M.

Para melhor compreender a qualidade de cada acerto na classificação nas operações de negociação, foi sumarizado na tabela 5.13 a média de lucro para as operações positivas e a média de prejuízo para as operações negativas. No apêndice estão as tabelas apêndice 1, apêndice 2 e apêndice 3 com todas as operações realizadas pelo algoritmo de negociação do modelo final.

Tabela 5.13: Sumarização das operações realizadas entre lucros e prejuízos

Ação	Lucro Médio (US\$)	Prejuízo Médio (US\$)	Operações Positivas	Operações Negativas
S&P 500	3.093,36	-2.107,35	126	38
MSFT	4.164,83	-2.679,84	107	71
XOM	2.895,53	-2.994,94	112	53
Média	3.384,57	-2.594,04	115	54

Capítulo 6

Conclusão e Trabalhos Futuros

6.1 Conclusão

O presente trabalho formulou um método para classificação de movimentos de ações na bolsa de valores. Como visto neste e outros trabalhos, a classificação de valores futuros para ações da bolsa de valores é uma tarefa complexa, devido aos conjuntos de dados serem extremamente ruidosos, sendo assim, o pré-processamento é a principal fase do processo de descoberta do conhecimento para este tipo de série temporal, sem aprimoramentos nesta fase o resultado é sempre próximo do aleatório. Para contornar este problema, o método desenvolvido otimiza duas camadas do pré-processamento: a transformação de atributos e a discretização. Além disto, foi feita também a seleção de atributos para que o algoritmo utilize apenas os atributos com maior contribuição na construção do classificador. Também foi escolhido um algoritmo do tipo *ensemble*, o *Random Forest*, que consegue lidar melhor com dados desta natureza, pois sua formulação voltada para amostragem, aleatoriedade de atributos e instâncias reduz o sobre ajuste da base de treinamento, além de reduzir a variância dos resultados.

Os atributos encontrados pelo algoritmo de otimização de janelas e seu processo de seleção de atributos, e em ordem de relevância, encontrados em pelo menos dois dos três experimentos foram:

- ADX, $n=6$ e 7
- Estocástico %K, $n=11, 14$ e 15
- William's %R, $n=10$ e 11

- Aroon Up, $n=25$
- RSI, $n=6$ e 7
- -DI, $n=6$ e 7

A proposta de adicionar os indicadores direcionais mostrou ser bastante efetiva, sendo o atributo ADX com melhor qualificação nos três experimentos, segundo o algoritmo de seleção de atributos ReliefF. Outra importante característica observada com o processo de otimização de janelas de tempo é que a maioria dos indicadores técnicos reduziram o tamanho da janela em relação a janela padrão, indicando que para a previsão do movimento no dia seguinte é mais eficiente com tamanhos de janelas mais reduzidos.

As operações realizadas pelo algoritmo de negociação, detalhadas nas tabelas apêndice 1, apêndice 2 e apêndice 3, apresentaram uma taxa relevante de lucro por operação acima das operações com prejuízo, conforme informações apresentadas na tabela 5.13. Além disso, pode-se perceber que o algoritmo realizou consideravelmente mais negociações com retorno positivo do que negativo. Ao total foram 345 negociações com retorno positivo e 162 negociações com retorno negativo, ou seja, o algoritmo de negociação obteve uma taxa de acerto de 68,04% das negociações. Sendo assim, uma taxa de acerto superior a precisão do classificador em prever o próximo dia, que em média ficou em 56,83% para os mesmos conjuntos de dados. Além disso, as operações com retorno positivo foram 76,64% maiores do que as operações com prejuízo. A taxa de acerto superior se deve principalmente ao classificador ter sido treinado com atributos de indicadores técnicos que em sua maioria indicam a tendência e não somente o movimento do dia seguinte, como ocorre com as negociações. A venda de uma ação pode ocorrer dias ou até semanas após a compra da ação, durante este período de tendência, vários dias podem ocorrer variações contrárias ao movimento classificado inicialmente, gerando erro para a taxa de acerto, mas não necessariamente para a taxa de acerto da negociação, pois se no final a venda obter lucro, a negociação foi positiva, gerando uma taxa de acerto para o algoritmo de negociação. Com isso, o algoritmo de negociação como validação se mostrou mais coeso com os atributos utilizados, além de ser o mais próximo da realidade deste mercado.

A adição do processo de otimização da discretização aumentou em torno de 4% a lucratividade do algoritmo além de ter aumentado as taxas de classificação em 5%, indicando assim que o processo para otimização de discretização é um passo importante.

6.2 Trabalhos Futuros

É possível propor outros trabalhos para a otimização de discretização dos atributos, visto que para um dos conjuntos de dados o processo de discretização não melhorou a previsão dos valores. Outra possível atividade para otimizar a discretização é a busca por tamanhos de categorias diferentes, como já estudado por Debie, Shafi, Merrick, & Lokan (2014) que utilizou deslocamento incremental de janelas, combinado com as outras técnicas proposta neste trabalho. Além disso, pode-se descobrir qual a quantidade de atributos e quais desses atributos o classificador pode obter a maior precisão.

Finalmente, outra possível melhoria é alterar o algoritmo de negociação para que possa levar em consideração mais de uma ação, podendo comprar a ação que tenha a maior probabilidade de subir no dia seguinte e trocar de ação caso uma delas passe a ter maior probabilidade.

Referências Bibliográficas

- Abu-Mostafa, Yaser S., e Amir F. Atiya. 1996. “Introduction to financial forecasting”. *Applied Intelligence* 6 (3): 205–13.
- Achelis, Steven B. 1995. *Technical Analysis from A to Z*. Probus Publishing.
- Amit, Yali, e Donald Geman. 1997. “Shape Quantization and Recognition with Randomized Trees”. *Neural Computation*.
- Bäck, Thomas, D.B Fogel, e Z Michalewicz. 2000a. *Evolutionary Computation 1: Basic Algorithms and Operators. Evolutionary Computation*. Taylor & Francis Group.
- Bäck, Thomas, D.B Fogel, e Zbigniew Michalewicz. 2000b. *Evolutionary Computation 2: Advanced Algorithms and Operators. Evolutionary Computation*.
- Bagheri, Ahmad, Hamed Mohammadi Peyhani, e Mohsen Akbari. 2014. “Financial forecasting using ANFIS networks with Quantum-behaved Particle Swarm Optimization”. *Expert Systems with Applications* 41 (14). Elsevier Ltd: 6235–50.
- Bollerslev, Tim, e Ian Domowitz. 1993. “Trading Patterns and Prices in the Interbank Foreign Exchange Market.pdf”. *Journal of Finance* 48 (4): 1421–43.
- Breiman, L, J H Friedman, R A Olshen, e C J Stone. 1984. *Classification and Regression Trees. The Wadsworth statistics probability series*. Vol. 19.
- Breiman, Leo. 1994. “Bagging predictors: Technical Report No. 421”. *Machine Learning*. Berkeley.
- Breiman, Leo. 2001. “Random Forests”. *Machine Learning*, 5–32.
- Brown, R. G. 1959. “Statistical forecasting for inventory control”. *McGraw-Hill, New York*, 1959.
- Brown, Robert Goodell. 1963. *Smoothing, Forecasting and Prediction of Discrete Time Series*. New Jersey: Englewood Cliffs.
- Chang, Pei-Chann, Chin-Yuan Fan, e Jun-Lin Lin. 2011. “Trend discovery in financial time series data using a case based fuzzy decision tree”. *Expert Systems with Applications* 38 (5). Elsevier Ltd: 6070–80.
- Chang, Pei-Chann, T. Warren Liao, Jyun-Jie Lin, e Chin-Yuan Fan. 2011. “A dynamic threshold decision system for stock trading signal detection”. *Applied Soft Computing* 11 (5). Elsevier B.V.: 3998–4010.
- De Toro, Francisco, Eduardo Ros, Sonia Mota, e Julio Ortega. 2006. “Evolutionary algorithms for multiobjective and multimodal optimization of diagnostic schemes”. *IEEE Transactions on Biomedical Engineering* 53 (2): 178–89.
- Debie, E, K Shafi, K Merrick, e C Lokan. 2014. “An online evolutionary rule learning algorithm with incremental attribute discretization”. In *Evolutionary Computation (CEC), 2014*

IEEE Congress on, 1116–23.

Faceli, Katti. 2011. *Inteligência artificial : uma abordagem de aprendizado máquina*. Rio de Janeiro: LTC.

Fama, Eugene. 1970. “Efficient Capital Markets: A Review of Theory and Empirical Work”. *Journal of Finance* 25 (2): 383–417.

Fang, Xiaozhao, Yong Xu, Xuelong Li, Zizhu Fan, Hong Liu, e Yan Chen. 2014. “Locality and similarity preserving embedding for feature selection”. *Neurocomputing* 128: 304–15.

Fu, Tak-chung. 2011. “A review on time series data mining”. *Engineering Applications of Artificial Intelligence* 24 (1). Elsevier: 164–81.

Gan, Min, Yu Cheng, Kai Liu, e Gang-lin Zhang. 2014. “Seasonal and trend time series forecasting based on a quasi-linear autoregressive model”. *Applied Soft Computing* 24 (novembro). Elsevier B.V.: 13–18.

Goldberg, David E, e Kalyanmoy Deb. 1991. “A comparative analysis of selection schemes used in genetic algorithms”. *Foundations of Genetic Algorithms* 1: 69–93.

Hall, Mark a. 1999. “Correlation-based Feature Selection for Machine Learning”. *Methodology*, 1–5.

Holt, Charles C. 1957. “Forecasting Trends and Seasonal by Exponentially Weighted Averages”. *Office of Naval Research Memorandum*.

John J Murphy. 1999. *Technical Analysis Of The Financial Markets*. New York Institute of Finance.

Kalecki, Michal. 1977. *Theory of Economic Dynamics - An Essay on Ciclical and Long-Run Changes in Capitalist Economy*. São Paulo: Hucitec.

Kara, Yakup, Melek Acar Boyacioglu, e Ömer Kaan Baykan. 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 (5). Elsevier Ltd: 5311–19.

Kim, Kyoung-jae. 2003. “Financial time series forecasting using support vector machines”. *Neurocomputing* 55 (1-2): 307–19.

Kim, Kyoung-jae, e Ingo Han. 2000. “Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index”. *Expert Systems with Applications* 19 (2): 125–32.

Kira, K, e L A Rendell. 1992. “The feature selection problem: Traditional methods and a new algorithm”. In *Proceedings Of The National Conference On Artificial Intelligence*, 35:129–34.

Kononenko, Igor. 1994. “Estimating attributes: Analysis and extensions of RELIEF”. *Machine Learning: ECML-94* 784: 171–82.

Liu, Huan, e Rudy Setiono. 1996. “A probabilistic approach to feature selection - a filter solution”. In *Proc 13th International Conference on Machine Learning*, 96:319–27.

Miao, K., F. Chen, e Z. G. Zhao. 2007. “Stock price forecast based on bacterial colony RBF neural network”. *Journal of QingDao University* 20: 50–54.

Mises, Ludwig Von. 1978. *On the Manipulation of Money and Credit*. Free Market Books. Free

Market Books.

- Mitchell, Melanie. 1998. "An Introduction to Genetic Algorithms (Complex Adaptive Systems)". *The MIT Press*, 221.
- Mitchell, T. 1990. *Machine Learning*. Vol. 4. McGraw-Hill Science/Engineering/Math.
- Molina, L C, Lluís Belanche, e Àngela Nebot. 2002. "Feature selection algorithms: A survey and experimental evaluation". *Data Mining 2002 ICDM 2002 Proceedings 2002 IEEE International Conference on* 4: 306–13.
- Patel, Jigar, Sahil Shah, Priyank Thakkar, e K Kotecha. 2015. "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques". *Expert Systems with Applications* 42 (1). Elsevier Ltd: 259–68.
- Quinlan, J R. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann San Mateo California. Vol. 1.
- Quinlan, J. R. 1986. "Induction of decision trees". *Machine Learning* 1 (1): 81–106.
- Ratnadip Adhikari, e R. K. Agrawal. 2013. *An Introductory Study on Time Series Modeling and Forecasting*. LAP LAMBERT Academic Publishing.
- Robnik-Šikonja, Marko, e Igor Kononenko. 2003. "Theoretical and Empirical Analysis of ReliefF and RReliefF". *Machine Learning* 53 (1-2): 23–69.
- Tin Kam Ho. 1995. "Random decision forests". *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1: 278–82.
- Trevor Hastie, Robert Tibshirani, e Jerome Friedman. 2008. "The Elements of Statistical Learning".
- Van Der Maaten, Laurens, Eric Postma, e Jaap Van Den Herik. 2009. "Dimensionality Reduction : A Comparative Review". *October*, 1–35.
- Vapnik, Vladimir N. 1998. "Statistical Learning Theory". *Interpreting*.
- Wang, Shuai, e Wei Shang. 2014. "Forecasting Direction of China Security Index 300 Movement with Least Squares Support Vector Machine". *Procedia Computer Science* 31. Elsevier Masson SAS: 869–74.
- Wilder Jr, J. Welles. 1978. *New Concepts in Technical Trading Systems*. Winston-Salem: Hunter Publishing Company.
- Witten, Ian H., Eibe Frank, e Mark A. Hall. 2011. *Data Mining Practical Machine Learning Tools and Techniques*. Third. Morgan Kaufmann.
- Żbikowski, Kamil. 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.

Apêndice

A seguir as tabelas com as negociações realizadas nos algoritmos para cada conjunto de dados.

Tabela Apêndice 1: Negociações executadas pelo algoritmo final na ação Microsoft

Data	Preço	Qtnd	Total (R\$)	Oper.	Lucro (R\$)	Data	Preço	Qtnd	Total (R\$)	Oper.	Lucro (R\$)
07/07/2010	21,07	4745	99987,64	Compra		27/08/2012	28,18	7158	201747,57	Compra	
08/07/2010	21,17	4745	100440,26	Vende	452,62	28/08/2012	28,13	7158	201353,14	Vende	-394,43
12/07/2010	21,53	4665	100445,90	Compra		18/09/2012	28,63	7032	201360,69	Compra	
13/07/2010	21,79	4665	101659,50	Vende	1213,60	20/09/2012	28,88	7032	203104,36	Vende	1743,66
14/07/2010	22,06	4608	101656,10	Compra		21/09/2012	28,64	7091	203115,28	Compra	
15/07/2010	22,12	4608	101935,81	Vende	279,71	27/09/2012	27,70	7091	196407,72	Vende	-6707,56
19/07/2010	21,88	4659	101932,77	Compra		01/10/2012	27,08	7252	196404,89	Compra	
21/07/2010	21,78	4659	101488,35	Vende	-444,41	03/10/2012	27,42	7252	198869,11	Vende	2464,22
23/07/2010	22,38	4535	101500,72	Compra		04/10/2012	27,58	7211	198870,59	Compra	
28/07/2010	22,50	4535	102051,29	Vende	550,57	08/10/2012	27,35	7211	197214,99	Vende	-1655,60
04/08/2010	22,31	4573	102033,98	Compra		11/10/2012	26,59	7417	197195,31	Compra	
06/08/2010	22,16	4573	101320,17	Vende	-713,80	12/10/2012	26,82	7417	198898,20	Vende	1702,89
12/08/2010	21,24	4771	101321,60	Compra		15/10/2012	27,10	7339	198895,90	Compra	
16/08/2010	21,25	4771	101362,97	Vende	41,37	18/10/2012	27,09	7339	198828,50	Vende	-67,40
17/08/2010	21,54	4706	101376,91	Compra		19/10/2012	26,30	7560	198844,93	Compra	
18/08/2010	21,64	4706	101828,21	Vende	451,30	22/10/2012	25,71	7560	194401,47	Vende	-4443,46
19/08/2010	21,31	4779	101824,59	Compra		24/10/2012	25,62	7587	194398,99	Compra	
20/08/2010	21,12	4779	100949,66	Vende	-874,93	26/10/2012	25,91	7587	196558,97	Vende	2159,98
23/08/2010	21,17	4769	100946,31	Compra		01/11/2012	27,11	7250	196550,47	Compra	
24/08/2010	20,96	4769	99948,48	Vende	-997,82	12/11/2012	25,92	7250	187894,78	Vende	-8655,69
25/08/2010	21,01	4757	99945,81	Compra		15/11/2012	24,69	7612	187902,59	Compra	
26/08/2010	20,77	4757	98784,61	Vende	-1161,20	23/11/2012	25,65	7612	195232,62	Vende	7330,04
30/08/2010	20,61	4793	98780,06	Compra		29/11/2012	24,95	7824	195236,70	Compra	
01/09/2010	20,84	4793	99866,48	Vende	1086,41	06/12/2012	24,75	7824	193642,92	Vende	-1593,78
03/09/2010	21,18	4716	99865,56	Compra		07/12/2012	24,50	7904	193646,92	Compra	
07/09/2010	20,89	4716	98508,79	Vende	-1356,76	10/12/2012	24,94	7904	197159,79	Vende	3512,88
09/09/2010	20,93	4707	98525,98	Compra		14/12/2012	24,82	7942	197151,70	Compra	
10/09/2010	20,79	4707	97869,41	Vende	-656,57	17/12/2012	25,09	7942	199284,27	Vende	2132,57
13/09/2010	21,89	4470	97851,76	Compra		28/12/2012	24,58	8107	199295,97	Compra	
16/09/2010	22,08	4470	98709,08	Vende	857,32	31/12/2012	24,73	8107	200497,00	Vende	1201,03
29/09/2010	21,36	4622	98721,20	Compra		04/01/2013	24,76	8097	200474,61	Compra	
05/10/2010	21,23	4622	98116,79	Vende	-604,41	07/01/2013	24,71	8097	200099,75	Vende	-374,86
08/10/2010	21,42	4580	98103,62	Compra		08/01/2013	24,58	8140	200107,22	Compra	
15/10/2010	22,27	4580	101976,66	Vende	3873,04	11/01/2013	24,84	8140	202217,58	Vende	2110,37
21/10/2010	22,16	4602	101985,06	Compra		15/01/2013	25,19	8026	202209,48	Compra	
22/10/2010	22,13	4602	101824,58	Vende	-160,48	16/01/2013	25,04	8026	200946,15	Vende	-1263,33

25/10/2010	21,96	4637	101830,92	Compra		18/01/2013	25,23	7965	200967,63	Compra	
10/11/2010	23,49	4637	108905,32	Vende	7074,40	28/01/2013	25,84	7965	205835,11	Vende	4867,48
12/11/2010	22,90	4755	108899,28	Compra		31/01/2013	25,42	8098	205823,03	Compra	
29/11/2010	22,20	4755	105564,38	Vende	-3334,90	08/02/2013	25,51	8098	206572,83	Vende	749,80
02/12/2010	23,59	4476	105573,66	Compra		15/02/2013	25,94	7965	206572,60	Compra	
14/12/2010	24,23	4476	108439,74	Vende	2866,08	19/02/2013	26,19	7965	208580,32	Vende	2007,71
20/12/2010	24,39	4445	108429,50	Compra		20/02/2013	26,02	8016	208568,82	Compra	
22/12/2010	24,73	4445	109911,11	Vende	1481,60	27/02/2013	25,96	8016	208119,79	Vende	-449,02
30/12/2010	24,43	4499	109904,61	Compra		28/02/2013	25,95	8019	208122,82	Compra	
10/01/2011	24,75	4499	111364,74	Vende	1460,13	04/03/2013	26,28	8019	210743,06	Vende	2620,25
13/01/2011	24,73	4504	111369,99	Compra		21/03/2013	26,24	8031	210758,54	Compra	
14/01/2011	24,82	4504	111804,56	Vende	434,57	22/03/2013	26,37	8031	211808,20	Vende	1049,66
19/01/2011	24,97	4477	111801,92	Compra		26/03/2013	26,29	8056	211790,65	Compra	
20/01/2011	24,87	4477	111330,69	Vende	-471,24	27/03/2013	26,49	8056	213370,07	Vende	1579,42
21/01/2011	24,58	4530	111337,40	Compra		12/04/2013	26,88	7939	213384,16	Compra	
24/01/2011	24,89	4530	112767,85	Vende	1430,45	17/04/2013	26,92	7939	213680,62	Vende	296,47
28/01/2011	24,34	4633	112771,67	Compra		19/04/2013	27,79	7688	213671,65	Compra	
01/02/2011	24,55	4633	113746,99	Vende	975,33	03/05/2013	31,27	7688	240371,65	Vende	26699,99
03/02/2011	24,25	4690	113747,72	Compra		09/05/2013	30,49	7884	240390,60	Compra	
07/02/2011	24,74	4690	116010,33	Vende	2262,62	16/05/2013	32,04	7884	252601,34	Vende	12210,74
10/02/2011	24,12	4809	116001,13	Compra		23/05/2013	32,11	7867	252574,39	Compra	
17/02/2011	24,01	4809	115456,24	Vende	-544,89	28/05/2013	32,92	7867	259008,92	Vende	6434,54
18/02/2011	23,88	4836	115464,42	Compra		31/05/2013	32,81	7894	259007,29	Compra	
22/02/2011	23,46	4836	113458,94	Vende	-2005,47	07/06/2013	33,53	7894	264721,75	Vende	5714,46
23/02/2011	23,46	4836	113458,94	Compra		13/06/2013	32,64	8110	264721,98	Compra	
03/03/2011	23,12	4836	111794,82	Vende	-1664,12	24/06/2013	31,70	8110	257097,50	Vende	-7624,48
16/03/2011	21,87	5111	111793,49	Compra		27/06/2013	32,55	7900	257124,56	Compra	
17/03/2011	21,86	5111	111748,40	Vende	-45,09	10/07/2013	32,62	7900	257718,74	Vende	594,17
30/03/2011	22,60	4945	111740,34	Compra		17/07/2013	33,60	7670	257714,80	Compra	
06/04/2011	23,07	4945	114096,43	Vende	2356,09	22/07/2013	30,09	7670	230818,41	Vende	-26896,39
14/04/2011	22,43	5087	114096,25	Compra		05/08/2013	29,80	7745	230818,23	Compra	
15/04/2011	22,38	5087	113871,83	Vende	-224,42	07/08/2013	30,14	7745	233439,52	Vende	2621,29
19/04/2011	22,19	5132	113882,95	Compra		12/08/2013	30,90	7554	233435,07	Compra	
26/04/2011	23,11	5132	118592,24	Vende	4709,28	13/08/2013	30,51	7554	230502,82	Vende	-2932,24
06/05/2011	22,83	5195	118581,27	Compra		14/08/2013	30,63	7526	230503,46	Compra	
09/05/2011	22,79	5195	118397,92	Vende	-183,35	20/08/2013	29,94	7526	225302,01	Vende	-5201,45
10/05/2011	22,65	5228	118411,95	Compra		26/08/2013	32,33	6968	225287,88	Compra	
31/05/2011	22,21	5228	116123,67	Vende	-2288,29	30/08/2013	31,62	6968	220340,13	Vende	-4947,75
01/06/2011	21,70	5352	116121,08	Compra		03/09/2013	30,18	7301	220363,48	Compra	
07/06/2011	21,37	5352	114362,38	Vende	-1758,69	10/09/2013	30,67	7301	223888,74	Vende	3525,27
08/06/2011	21,26	5378	114344,80	Compra		11/09/2013	31,00	7222	223859,31	Compra	
14/06/2011	21,51	5378	115682,16	Vende	1337,36	18/09/2013	31,55	7222	227825,03	Vende	3965,72
15/06/2011	21,08	5487	115687,69	Compra		20/09/2013	31,04	7339	227833,35	Compra	
21/06/2011	21,99	5487	120658,26	Vende	4970,57	27/09/2013	31,50	7339	231168,50	Vende	3335,16
22/06/2011	21,89	5512	120669,53	Compra		01/10/2013	31,79	7271	231160,61	Compra	
23/06/2011	21,87	5512	120571,62	Vende	-97,91	04/10/2013	32,08	7271	233225,77	Vende	2065,16
29/06/2011	22,75	5299	120571,46	Compra		07/10/2013	31,53	7398	233237,04	Compra	
13/07/2011	23,65	5299	125324,66	Vende	4753,20	09/10/2013	31,31	7398	231626,10	Vende	-1610,94
27/07/2011	24,27	5163	125317,94	Compra		14/10/2013	32,62	7102	231637,50	Compra	
01/08/2011	24,22	5163	125042,82	Vende	-275,12	16/10/2013	32,80	7102	232915,03	Vende	1277,53
08/08/2011	21,74	5751	125033,46	Compra		30/10/2013	33,65	6922	232909,93	Compra	
09/08/2011	22,72	5751	130651,79	Vende	5618,34	20/11/2013	35,37	6922	244845,17	Vende	11935,24

10/08/2011	21,49	6079	130652,88	Compra		21/11/2013	35,68	6863	244853,22	Compra	
11/08/2011	22,37	6079	135997,77	Vende	5344,89	29/11/2013	36,37	6863	249632,44	Vende	4779,22
18/08/2011	22,05	6169	136015,35	Compra		05/12/2013	36,25	6886	249615,08	Compra	
26/08/2011	22,57	6169	139213,11	Vende	3197,76	06/12/2013	36,59	6886	251979,86	Vende	2364,78
29/08/2011	23,09	6028	139209,78	Compra		13/12/2013	35,00	7200	251999,92	Compra	
30/08/2011	23,44	6028	141310,85	Vende	2101,07	16/12/2013	35,19	7200	253373,59	Vende	1373,67
08/09/2011	23,43	6030	141303,84	Compra		19/12/2013	34,58	7327	253369,54	Compra	
14/09/2011	23,68	6030	142812,81	Vende	1508,97	20/12/2013	35,10	7327	257213,76	Vende	3844,22
19/09/2011	24,32	5873	142821,15	Compra		06/01/2014	34,47	7463	257218,16	Compra	
20/09/2011	24,11	5873	141613,91	Vende	-1207,24	07/01/2014	34,73	7463	259211,54	Vende	1993,37
30/09/2011	22,24	6366	141610,52	Compra		08/01/2014	34,11	7598	259189,25	Compra	
06/10/2011	23,54	6366	149860,23	Vende	8249,71	10/01/2014	34,38	7598	261218,72	Vende	2029,46
07/10/2011	23,46	6388	149864,30	Compra		13/01/2014	33,37	7828	261210,61	Compra	
10/10/2011	24,08	6388	153803,59	Vende	3939,29	14/01/2014	34,13	7828	267184,55	Vende	5973,94
12/10/2011	24,09	6383	153797,29	Compra		16/01/2014	35,19	7593	267203,57	Compra	
13/10/2011	24,29	6383	155052,33	Vende	1255,03	21/01/2014	34,50	7593	261988,42	Vende	-5215,15
25/10/2011	23,96	6471	155050,15	Compra		27/01/2014	34,37	7622	261971,11	Compra	
26/10/2011	23,76	6471	153777,83	Vende	-1272,32	30/01/2014	35,16	7622	268005,98	Vende	6034,87
27/10/2011	24,35	6314	153771,23	Compra		04/02/2014	34,68	7729	268008,09	Compra	
31/10/2011	23,80	6314	150272,57	Vende	-3498,66	19/02/2014	36,05	7729	278634,60	Vende	10626,51
01/11/2011	23,23	6470	150284,62	Compra		20/02/2014	36,28	7680	278639,62	Compra	
03/11/2011	23,71	6470	153407,12	Vende	3122,50	26/02/2014	36,01	7680	276572,90	Vende	-2066,72
04/11/2011	23,46	6539	153406,80	Compra		06/03/2014	36,67	7543	276568,91	Compra	
07/11/2011	23,95	6539	156621,04	Vende	3214,24	10/03/2014	36,35	7543	274176,55	Vende	-2392,35
09/11/2011	23,42	6689	156626,95	Compra		17/03/2014	36,57	7498	274198,31	Compra	
10/11/2011	23,49	6689	157105,20	Vende	478,25	19/03/2014	37,74	7498	282989,97	Vende	8791,65
14/11/2011	23,92	6569	157104,76	Compra		21/03/2014	38,60	7331	282957,76	Compra	
15/11/2011	24,08	6569	158169,47	Vende	1064,71	25/03/2014	38,77	7331	284226,00	Vende	1268,24
16/11/2011	23,47	6737	158150,15	Compra		26/03/2014	38,24	7433	284251,51	Compra	
18/11/2011	22,78	6737	153479,04	Vende	-4671,11	31/03/2014	39,40	7433	292824,06	Vende	8572,55
22/11/2011	22,32	6876	153488,00	Compra		02/04/2014	39,74	7368	292812,63	Compra	
30/11/2011	23,03	6876	158379,30	Vende	4891,30	03/04/2014	39,41	7368	290404,98	Vende	-2407,65
15/12/2011	23,02	6881	158370,54	Compra		04/04/2014	38,32	7579	290417,53	Compra	
16/12/2011	23,41	6881	161096,80	Vende	2726,26	07/04/2014	38,25	7579	289907,64	Vende	-509,88
21/12/2011	23,20	6945	161094,28	Compra		10/04/2014	37,83	7664	289918,07	Compra	
23/12/2011	23,44	6945	162782,77	Vende	1688,49	15/04/2014	38,20	7664	292790,73	Vende	2872,66
28/12/2011	23,25	7002	162794,74	Compra		02/05/2014	38,15	7675	292768,37	Compra	
04/01/2012	24,67	7002	172756,61	Vende	9961,88	07/05/2014	37,89	7675	290776,75	Vende	-1991,62
09/01/2012	24,98	6916	172752,14	Compra		09/05/2014	38,00	7652	290787,90	Compra	
10/01/2012	25,07	6916	173374,90	Vende	622,76	12/05/2014	38,41	7652	293950,23	Vende	3162,33
11/01/2012	24,96	6946	173376,41	Compra		27/05/2014	38,90	7557	293957,92	Compra	
20/01/2012	26,75	6946	185822,98	Vende	12446,57	28/05/2014	38,72	7557	292641,36	Vende	-1316,57
02/02/2012	26,97	6890	185813,85	Compra		03/06/2014	39,00	7504	292622,59	Compra	
06/02/2012	27,19	6890	187364,88	Vende	1551,04	04/06/2014	39,02	7504	292840,47	Vende	217,88
13/02/2012	27,54	6805	187381,89	Compra		06/06/2014	40,15	7294	292834,51	Compra	
14/02/2012	27,42	6805	186580,07	Vende	-801,83	09/06/2014	39,94	7294	291351,99	Vende	-1482,52
17/02/2012	28,32	6587	186573,27	Compra		02/07/2014	40,55	7185	291379,21	Compra	
28/02/2012	28,89	6587	190274,89	Vende	3701,62	07/07/2014	40,64	7185	292005,08	Vende	625,87
02/03/2012	29,08	6544	190278,37	Compra		08/07/2014	40,44	7221	292000,45	Compra	
05/03/2012	28,82	6544	188617,57	Vende	-1660,80	09/07/2014	40,33	7221	291231,65	Vende	-768,80
06/03/2012	28,61	6594	188624,31	Compra		10/07/2014	40,35	7217	291210,03	Compra	
13/03/2012	29,61	6594	195258,43	Vende	6634,12	11/07/2014	40,74	7217	294004,10	Vende	2794,06

16/03/2012	29,55	6608	195253,74	Compra		15/07/2014	41,09	7156	294012,49	Compra	
20/03/2012	29,00	6608	191600,22	Vende	-3653,52	21/07/2014	43,40	7156	310565,84	Vende	16553,35
30/03/2012	29,24	6552	191579,91	Compra		29/07/2014	42,48	7311	310570,43	Compra	
04/04/2012	28,29	6552	185344,37	Vende	-6235,54	31/07/2014	41,77	7311	305404,87	Vende	-5165,56
05/04/2012	28,57	6488	185356,92	Compra		01/08/2014	41,48	7362	305397,68	Compra	
16/04/2012	28,17	6488	182769,44	Vende	-2587,47	07/08/2014	41,84	7362	308034,10	Vende	2636,41
17/04/2012	28,50	6414	182777,71	Compra		14/08/2014	42,85	7189	308031,94	Compra	
20/04/2012	29,38	6414	188474,97	Vende	5697,25	29/08/2014	44,25	7189	318077,59	Vende	10045,64
23/04/2012	29,11	6474	188477,69	Compra		03/09/2014	43,79	7264	318070,93	Compra	
01/05/2012	29,01	6474	187832,22	Vende	-645,47	04/09/2014	44,08	7264	320193,28	Vende	2122,35
04/05/2012	28,08	6689	187825,41	Compra		05/09/2014	44,71	7161	320186,35	Compra	
11/05/2012	28,24	6689	188916,71	Vende	1091,30	12/09/2014	45,48	7161	325695,99	Vende	5509,64
14/05/2012	27,81	6794	188926,40	Compra		16/09/2014	45,54	7152	325704,55	Compra	
15/05/2012	27,56	6794	187252,82	Vende	-1673,58	17/09/2014	45,31	7152	324032,87	Vende	-1671,69
16/05/2012	27,28	6864	187240,84	Compra		25/09/2014	44,84	7227	324052,39	Compra	
21/05/2012	27,14	6864	186301,50	Vende	-939,34	26/09/2014	45,20	7227	326656,62	Vende	2604,23
29/05/2012	26,97	6908	186298,29	Compra		01/10/2014	44,70	7307	326643,22	Compra	
30/05/2012	26,77	6908	184911,77	Vende	-1386,52	03/10/2014	44,89	7307	327995,33	Vende	1352,11
01/06/2012	25,96	7124	184909,11	Compra		08/10/2014	45,56	7199	327985,18	Compra	
06/06/2012	26,78	7124	190758,61	Vende	5849,50	09/10/2014	44,65	7199	321464,74	Vende	-6520,44
11/06/2012	26,37	7235	190760,51	Compra		13/10/2014	42,51	7562	321471,72	Compra	
15/06/2012	27,39	7235	198153,32	Vende	7392,80	21/10/2014	43,71	7562	330530,38	Vende	9058,66
21/06/2012	27,50	7206	198147,97	Compra		22/10/2014	43,22	7647	330521,90	Compra	
29/06/2012	27,91	7206	201106,38	Vende	2958,42	10/11/2014	47,61	7647	364110,30	Vende	33588,40
03/07/2012	28,06	7166	201101,47	Compra		11/11/2014	47,60	7650	364104,13	Compra	
05/07/2012	28,01	7166	200709,21	Vende	-392,26	12/11/2014	47,51	7650	363433,59	Vende	-670,54
06/07/2012	27,54	7288	200735,23	Compra		17/11/2014	48,17	7545	363442,07	Compra	
10/07/2012	27,13	7288	197743,15	Vende	-2992,08	20/11/2014	47,73	7545	360114,52	Vende	-3327,55
11/07/2012	26,73	7397	197731,26	Compra		24/11/2014	46,64	7721	360115,38	Compra	
16/07/2012	26,86	7397	198676,06	Vende	944,80	03/12/2014	47,12	7721	363823,25	Vende	3707,86
17/07/2012	27,06	7342	198672,44	Compra		04/12/2014	47,87	7601	363830,25	Compra	
23/07/2012	26,71	7342	196127,09	Vende	-2545,35	10/12/2014	45,96	7601	349378,36	Vende	-14451,89
24/07/2012	26,59	7375	196133,91	Compra		12/12/2014	46,01	7593	349382,72	Compra	
27/07/2012	27,15	7375	200238,26	Vende	4104,35	23/12/2014	47,48	7593	360545,11	Vende	11162,39
01/08/2012	26,83	7463	200244,49	Compra		24/12/2014	47,18	7642	360550,03	Compra	
07/08/2012	27,61	7463	206031,91	Vende	5787,41	26/12/2014	46,93	7642	358602,74	Vende	-1947,29
17/08/2012	28,38	7260	206022,58	Compra		29/12/2014	46,50	7711	358590,96	Compra	
23/08/2012	27,79	7260	201755,45	Vende	-4267,13	30/12/2014	46,08	7711	355341,36	Vende	-3249,61

Tabela Apêndice 2: Negociações executadas pelo algoritmo final na ação Exxon Mobil

Data	Preço	Qtnd	Total (R\$)	Oper.	Lucro (R\$)	Data	Preço	Qtnd	Total (R\$)	Oper.	Lucro (R\$)
06/07/2010	49,86	2005	99973,80	Compra		29/08/2012	80,64	2573	207491,56	Vende	-354,09
08/07/2010	51,03	2005	102322,65	Vende	2348,85	31/08/2012	80,09	2591	207516,89	Compra	
19/07/2010	50,70	2018	102320,65	Compra		17/09/2012	84,32	2591	218475,11	Vende	10958,23
21/07/2010	50,48	2018	101865,34	Vende	-455,31	18/09/2012	84,08	2598	218445,65	Compra	
22/07/2010	51,53	1977	101871,59	Compra		24/09/2012	84,29	2598	218993,84	Vende	548,19
27/07/2010	52,77	1977	104324,88	Vende	2453,29	27/09/2012	84,22	2600	218971,61	Compra	
28/07/2010	52,86	1974	104337,87	Compra		08/10/2012	85,03	2600	221070,68	Vende	2099,07

02/08/2010	53,75	1974	106102,24	Vende	1764,37	11/10/2012	83,64	2644	221149,10	Compra	
05/08/2010	54,42	1949	106060,79	Compra		22/10/2012	84,58	2644	223623,30	Vende	2474,20
13/08/2010	52,36	1949	102050,23	Vende	-4010,56	23/10/2012	82,73	2702	223546,23	Compra	
16/08/2010	52,33	1950	102051,46	Compra		24/10/2012	82,46	2702	222802,56	Vende	-743,67
23/08/2010	52,00	1950	101403,84	Vende	-647,62	26/10/2012	83,14	2680	222807,93	Compra	
25/08/2010	51,49	1970	101428,05	Compra		12/11/2012	80,61	2680	216038,38	Vende	-6769,54
01/09/2010	53,23	1970	104871,54	Vende	3443,49	13/11/2012	79,83	2707	216090,72	Compra	
02/09/2010	53,37	1965	104862,98	Compra		14/11/2012	79,46	2707	215091,11	Vende	-999,61
07/09/2010	52,92	1965	103987,11	Vende	-875,87	15/11/2012	79,52	2705	215106,99	Compra	
08/09/2010	53,09	1959	104012,02	Compra		26/11/2012	81,81	2705	221300,01	Vende	6193,01
15/09/2010	53,31	1959	104440,06	Vende	428,03	27/11/2012	80,64	2744	221273,50	Compra	
17/09/2010	53,12	1966	104435,23	Compra		04/12/2012	80,49	2744	220868,20	Vende	-405,30
20/09/2010	53,79	1966	105758,28	Vende	1323,06	10/12/2012	81,62	2706	220857,22	Compra	
23/09/2010	53,44	1979	105765,76	Compra		11/12/2012	82,15	2706	222306,10	Vende	1448,89
27/09/2010	53,93	1979	106734,34	Vende	968,58	12/12/2012	82,58	2692	222299,14	Compra	
30/09/2010	54,00	1976	106710,70	Compra		14/12/2012	81,31	2692	218894,46	Vende	-3404,68
01/10/2010	54,66	1976	108005,94	Vende	1295,24	17/12/2012	82,04	2668	218888,74	Compra	
04/10/2010	54,35	1987	107999,37	Compra		20/12/2012	82,06	2668	218937,99	Vende	49,25
05/10/2010	55,29	1987	109857,54	Vende	1858,16	21/12/2012	80,53	2719	218956,33	Compra	
11/10/2010	56,42	1947	109841,14	Compra		31/12/2012	79,90	2719	217249,45	Vende	-1706,87
14/10/2010	57,07	1947	111117,38	Vende	1276,23	15/01/2013	82,65	2628	217208,26	Compra	
15/10/2010	56,97	1951	111158,09	Compra		24/01/2013	84,33	2628	221623,75	Vende	4415,49
18/10/2010	57,93	1951	113016,69	Vende	1858,60	28/01/2013	84,11	2635	221630,27	Compra	
19/10/2010	56,91	1985	112973,80	Compra		29/01/2013	84,71	2635	223211,43	Vende	1581,17
21/10/2010	57,96	1985	115055,62	Vende	2081,82	01/02/2013	83,12	2686	223266,67	Compra	
25/10/2010	57,86	1989	115078,87	Compra		06/02/2013	82,89	2686	222646,76	Vende	-619,91
27/10/2010	57,39	1989	114157,54	Vende	-921,32	07/02/2013	81,99	2715	222603,87	Compra	
29/10/2010	58,11	1965	114188,32	Compra		11/02/2013	82,02	2715	222679,54	Vende	75,67
02/11/2010	59,29	1965	116506,78	Vende	2318,45	12/02/2013	82,19	2710	222722,65	Compra	
09/11/2010	62,12	1875	116471,39	Compra		13/02/2013	82,38	2710	223251,38	Vende	528,73
10/11/2010	62,56	1875	117295,91	Vende	824,52	22/02/2013	82,87	2694	223259,83	Compra	
16/11/2010	60,63	1935	117322,43	Compra		05/03/2013	83,25	2694	224286,03	Vende	1026,20
22/11/2010	61,73	1935	119449,69	Vende	2127,26	11/03/2013	82,84	2707	224236,59	Compra	
23/11/2010	60,67	1969	119453,18	Compra		12/03/2013	82,84	2707	224236,59	Vende	0,00
29/11/2010	61,08	1969	120267,07	Vende	813,89	18/03/2013	82,47	2719	224245,41	Compra	
07/12/2010	62,85	1913	120228,32	Compra		27/03/2013	84,16	2719	228817,74	Vende	4572,32
09/12/2010	63,32	1913	121136,85	Vende	908,53	28/03/2013	83,72	2733	228802,51	Compra	
10/12/2010	63,48	1908	121122,28	Compra		01/04/2013	84,33	2733	230478,33	Vende	1675,83
04/01/2011	65,87	1908	125686,60	Vende	4564,32	03/04/2013	83,55	2759	230517,79	Compra	
05/01/2011	65,70	1914	125745,17	Compra		15/04/2013	80,36	2759	221700,02	Vende	-8817,77
13/01/2011	67,47	1914	129128,68	Vende	3383,51	30/04/2013	82,68	2681	221659,41	Compra	
14/01/2011	68,46	1886	129113,99	Compra		02/05/2013	82,34	2681	220762,70	Vende	-896,70
18/01/2011	69,22	1886	130557,07	Vende	1443,08	09/05/2013	85,16	2593	220815,32	Compra	
20/01/2011	68,38	1909	130537,45	Compra		14/05/2013	85,23	2593	221009,38	Vende	194,06
21/01/2011	69,46	1909	132602,55	Vende	2065,10	15/05/2013	85,35	2589	220983,30	Compra	
25/01/2011	69,20	1916	132583,25	Compra		17/05/2013	85,84	2589	222242,75	Vende	1259,45
31/01/2011	70,96	1916	135953,44	Vende	3370,19	31/05/2013	84,63	2626	222249,84	Compra	
02/02/2011	73,36	1854	136005,56	Compra		04/06/2013	84,93	2626	223035,96	Vende	786,12
07/03/2011	74,90	1854	138869,62	Vende	2864,06	05/06/2013	83,87	2659	223003,04	Compra	

11/03/2011	72,60	1912	138818,84	Compra		07/06/2013	85,55	2659	227480,50	Vende	4477,46
16/03/2011	70,12	1912	134068,70	Vende	-4750,14	10/06/2013	85,24	2669	227512,07	Compra	
24/03/2011	73,14	1833	134071,69	Compra		13/06/2013	85,44	2669	228036,40	Vende	524,33
25/03/2011	73,93	1833	135514,01	Vende	1442,33	19/06/2013	85,13	2679	228063,74	Compra	
06/04/2011	75,31	1800	135556,92	Compra		28/06/2013	84,52	2679	226434,71	Vende	-1629,03
07/04/2011	75,82	1800	136479,95	Vende	923,02	03/07/2013	84,84	2669	226438,43	Compra	
15/04/2011	74,52	1831	136450,76	Compra		05/07/2013	85,66	2669	228635,64	Vende	2197,21
18/04/2011	73,47	1831	134524,36	Vende	-1926,41	10/07/2013	86,83	2633	228630,71	Compra	
29/04/2011	77,78	1729	134490,16	Compra		18/07/2013	88,29	2633	232473,24	Vende	3842,53
03/05/2011	75,70	1729	130882,56	Vende	-3607,60	19/07/2013	89,03	2611	232460,45	Compra	
16/05/2011	71,34	1835	130916,59	Compra		23/07/2013	89,06	2611	232533,73	Vende	73,28
20/05/2011	72,53	1835	133086,57	Vende	2169,98	24/07/2013	88,86	2616	232465,10	Compra	
23/05/2011	71,73	1855	133052,69	Compra		25/07/2013	88,84	2616	232416,16	Vende	-48,94
27/05/2011	73,47	1855	136285,41	Vende	3232,72	09/08/2013	85,45	2720	232437,11	Compra	
06/06/2011	71,39	1910	136352,32	Compra		14/08/2013	83,79	2720	227902,12	Vende	-4534,99
14/06/2011	71,47	1910	136505,16	Vende	152,84	15/08/2013	83,04	2745	227954,13	Compra	
15/06/2011	69,94	1951	136451,70	Compra		30/08/2013	82,10	2745	225368,45	Vende	-2585,69
21/06/2011	71,64	1951	139764,97	Vende	3313,27	05/09/2013	82,28	2739	225366,03	Compra	
24/06/2011	68,27	2048	139812,42	Compra		09/09/2013	82,93	2739	227146,26	Vende	1780,23
27/06/2011	69,28	2048	141888,30	Vende	2075,88	11/09/2013	83,68	2714	227118,18	Compra	
28/06/2011	70,80	2004	141886,84	Compra		12/09/2013	82,87	2714	224919,62	Vende	-2198,56
01/07/2011	72,92	2004	146127,60	Vende	4240,76	13/09/2013	83,27	2701	224910,84	Compra	
05/07/2011	72,55	2014	146122,58	Compra		24/09/2013	82,29	2701	222264,82	Vende	-2646,01
06/07/2011	72,53	2014	146068,86	Vende	-53,72	01/10/2013	81,01	2744	222288,03	Compra	
13/07/2011	73,34	1991	146011,69	Compra		04/10/2013	81,31	2744	223115,15	Vende	827,12
15/07/2011	73,80	1991	146932,23	Vende	920,53	07/10/2013	80,91	2757	223081,45	Compra	
20/07/2011	74,06	1984	146944,86	Compra		29/10/2013	83,77	2757	230950,33	Vende	7868,88
21/07/2011	75,59	1984	149979,00	Vende	3034,15	30/10/2013	83,66	2761	230973,30	Compra	
25/07/2011	75,19	1995	150012,32	Compra		14/11/2013	88,41	2761	244092,28	Vende	13118,98
28/07/2011	72,43	1995	144495,73	Vende	-5516,59	18/11/2013	90,52	2696	244047,49	Compra	
10/08/2011	60,89	2373	144485,59	Compra		19/11/2013	90,49	2696	243970,79	Vende	-76,70
11/08/2011	64,06	2373	152025,26	Vende	7539,68	22/11/2013	90,10	2708	244003,76	Compra	
15/08/2011	66,49	2286	151996,26	Compra		26/11/2013	89,40	2708	242103,29	Vende	-1900,47
18/08/2011	63,49	2286	145142,22	Vende	-6854,05	27/11/2013	88,96	2722	242141,66	Compra	
24/08/2011	65,82	2206	145196,27	Compra		02/12/2013	88,69	2722	241418,83	Vende	-722,83
31/08/2011	66,25	2206	146143,97	Vende	947,70	05/12/2013	89,27	2704	241386,66	Compra	
02/09/2011	64,57	2263	146112,38	Compra		11/12/2013	89,39	2704	241720,05	Vende	333,38
14/09/2011	65,01	2263	147125,08	Vende	1012,70	12/12/2013	90,44	2673	241737,34	Compra	
16/09/2011	66,72	2205	147123,68	Compra		18/12/2013	94,40	2673	252333,63	Vende	10596,29
19/09/2011	65,96	2205	145446,20	Vende	-1677,48	20/12/2013	93,59	2696	252305,99	Compra	
22/09/2011	61,97	2347	145444,22	Compra		27/12/2013	96,27	2696	259541,77	Vende	7235,78
29/09/2011	66,12	2347	155190,91	Vende	9746,69	31/12/2013	95,98	2704	259516,95	Compra	
14/10/2011	69,91	2220	155197,94	Compra		03/01/2014	94,37	2704	255183,13	Vende	-4333,82
21/10/2011	71,72	2220	159211,50	Vende	4013,56	06/01/2014	94,51	2700	255189,73	Compra	
25/10/2011	71,10	2239	159191,42	Compra		07/01/2014	95,85	2700	258800,17	Vende	3610,44
26/10/2011	72,56	2239	162457,81	Vende	3266,39	15/01/2014	93,68	2763	258838,24	Compra	
31/10/2011	69,89	2325	162496,75	Compra		16/01/2014	93,83	2763	259257,50	Vende	419,27
01/11/2011	67,97	2325	158022,85	Vende	-4473,90	23/01/2014	91,96	2819	259245,37	Compra	
02/11/2011	69,25	2282	158020,92	Compra		28/01/2014	90,71	2819	255716,41	Vende	-3528,97

03/11/2011	70,58	2282	161064,10	Vende	3043,18	29/01/2014	90,20	2835	255715,93	Compra	
04/11/2011	70,28	2292	161072,43	Compra		07/02/2014	86,51	2835	245261,30	Vende	-10454,63
11/11/2011	71,78	2292	164508,47	Vende	3436,04	10/02/2014	85,50	2869	245298,13	Compra	
23/11/2011	67,15	2450	164510,96	Compra		14/02/2014	89,88	2869	257875,43	Vende	12577,30
28/11/2011	68,28	2450	167290,29	Vende	2779,34	19/02/2014	89,73	2874	257885,65	Compra	
29/11/2011	69,26	2415	167270,45	Compra		24/02/2014	92,11	2874	264720,53	Vende	6834,88
02/12/2011	71,84	2415	173489,01	Vende	6218,56	27/02/2014	91,54	2892	264721,20	Compra	
08/12/2011	71,91	2413	173519,14	Compra		18/03/2014	90,46	2892	261600,01	Vende	-3121,19
13/12/2011	72,50	2413	174952,99	Vende	1433,85	25/03/2014	90,65	2885	261517,92	Compra	
14/12/2011	71,52	2446	174945,21	Compra		09/04/2014	92,96	2885	268186,05	Vende	6668,14
16/12/2011	72,17	2446	176530,82	Vende	1585,61	14/04/2014	93,46	2870	268244,46	Compra	
22/12/2011	75,89	2326	176519,27	Compra		16/04/2014	95,45	2870	273945,96	Vende	5701,50
27/12/2011	76,78	2326	178592,51	Vende	2073,24	17/04/2014	95,91	2856	273918,93	Compra	
28/12/2011	75,79	2356	178562,63	Compra		07/05/2014	98,48	2856	281256,54	Vende	7337,61
29/12/2011	76,77	2356	180874,73	Vende	2312,10	08/05/2014	97,72	2878	281251,57	Compra	
04/01/2012	77,45	2336	180916,68	Compra		12/05/2014	98,30	2878	282912,04	Vende	1660,46
10/01/2012	77,18	2336	180285,73	Vende	-630,95	16/05/2014	96,87	2921	282953,95	Compra	
12/01/2012	76,29	2363	180284,56	Compra		19/05/2014	96,99	2921	283319,10	Vende	365,15
17/01/2012	77,15	2363	182305,69	Vende	2021,14	20/05/2014	96,80	2926	283241,34	Compra	
24/01/2012	78,49	2322	182257,51	Compra		21/05/2014	98,11	2926	287067,79	Vende	3826,44
25/01/2012	78,53	2322	182341,14	Vende	83,62	28/05/2014	97,18	2955	287156,75	Compra	
27/01/2012	77,28	2360	182371,71	Compra		03/06/2014	96,53	2955	285252,98	Vende	-1903,77
13/02/2012	76,42	2360	180357,52	Vende	-2014,19	06/06/2014	97,70	2919	285174,09	Compra	
16/02/2012	77,32	2333	180384,99	Compra		11/06/2014	98,03	2919	286156,47	Vende	982,39
22/02/2012	78,69	2333	183574,08	Vende	3189,09	12/06/2014	97,75	2928	286222,29	Compra	
02/03/2012	78,15	2348	183500,29	Compra		13/06/2014	98,71	2928	289009,62	Vende	2787,32
07/03/2012	77,70	2348	182437,51	Vende	-1062,78	17/06/2014	98,48	2934	288952,95	Compra	
08/03/2012	76,79	2376	182462,17	Compra		20/06/2014	99,84	2934	292930,93	Vende	3977,98
15/03/2012	77,93	2376	185172,31	Vende	2710,14	25/06/2014	98,22	2983	292975,54	Compra	
16/03/2012	78,25	2367	185220,88	Compra		08/07/2014	98,88	2983	294954,73	Vende	1979,18
19/03/2012	78,75	2367	186399,39	Vende	1178,51	22/07/2014	99,56	2962	294900,48	Compra	
21/03/2012	77,86	2393	186323,90	Compra		24/07/2014	100,27	2962	297008,13	Vende	2107,65
26/03/2012	78,79	2393	188533,53	Vende	2209,62	25/07/2014	99,22	2994	297050,02	Compra	
05/04/2012	76,78	2456	188583,44	Compra		18/08/2014	96,36	2994	288508,95	Vende	-8541,06
10/04/2012	74,39	2456	182691,59	Vende	-5891,85	26/08/2014	96,48	2990	288470,93	Compra	
11/04/2012	74,87	2440	182672,11	Compra		02/09/2014	95,37	2990	285141,52	Vende	-3329,41
17/04/2012	77,36	2440	188746,45	Vende	6074,34	10/09/2014	93,74	3042	285152,09	Compra	
08/05/2012	76,06	2482	188782,33	Compra		16/09/2014	94,34	3042	286978,30	Vende	1826,21
11/05/2012	75,75	2482	188001,51	Vende	-780,83	19/09/2014	94,04	3052	287005,59	Compra	
15/05/2012	74,55	2521	187945,36	Compra		23/09/2014	92,98	3052	283784,45	Vende	-3221,14
16/05/2012	74,90	2521	188818,55	Vende	873,20	25/09/2014	91,26	3110	283817,32	Compra	
18/05/2012	74,26	2543	188843,75	Compra		26/09/2014	92,40	3110	287370,68	Vende	3553,36
12/06/2012	74,07	2543	188356,98	Vende	-486,77	30/09/2014	91,07	3155	287313,03	Compra	
15/06/2012	75,86	2483	188348,85	Compra		02/10/2014	90,34	3155	285021,86	Vende	-2291,17
18/06/2012	75,76	2483	188099,89	Vende	-248,96	09/10/2014	88,91	3206	285034,84	Compra	
25/06/2012	74,05	2540	188088,46	Compra		14/10/2014	87,68	3206	281092,42	Vende	-3942,42
28/06/2012	75,75	2540	192394,77	Vende	4306,31	17/10/2014	88,32	3183	281109,96	Compra	
10/07/2012	75,76	2540	192417,93	Compra		20/10/2014	88,86	3183	282835,88	Vende	1725,92
13/07/2012	77,91	2540	197881,85	Vende	5463,92	27/10/2014	90,74	3117	282826,36	Compra	

24/07/2012	77,10	2567	197902,87	Compra		05/11/2014	92,05	3117	286930,98	Vende	4104,62
26/07/2012	78,86	2567	202442,13	Vende	4539,26	06/11/2014	93,21	3078	286887,50	Compra	
31/07/2012	79,16	2557	202422,64	Compra		07/11/2014	94,20	3078	289949,37	Vende	3061,87
01/08/2012	79,22	2557	202562,49	Vende	139,86	13/11/2014	92,32	3141	289971,86	Compra	
02/08/2012	78,28	2588	202588,52	Compra		18/11/2014	92,52	3141	290615,15	Vende	643,29
07/08/2012	80,14	2588	207400,82	Vende	4812,31	25/11/2014	92,44	3144	290616,75	Compra	
08/08/2012	80,47	2577	207364,91	Compra		04/12/2014	92,04	3144	289359,61	Vende	-1257,14
13/08/2012	80,86	2577	208381,54	Vende	1016,62	05/12/2014	91,50	3163	289411,66	Compra	
14/08/2012	80,92	2575	208361,55	Compra		10/12/2014	86,48	3163	273525,17	Vende	-15886,49
17/08/2012	81,10	2575	208834,03	Vende	472,49	11/12/2014	86,99	3144	273507,21	Compra	
22/08/2012	80,49	2595	208860,96	Compra		12/12/2014	84,46	3144	265535,03	Vende	-7972,18
23/08/2012	80,10	2595	207861,05	Vende	-999,92	29/12/2014	90,77	2925	265495,39	Compra	
24/08/2012	80,78	2573	207845,65	Compra							

Tabela Apêndice 3: Negociações executadas pelo algoritmo final no índice S&P 500

Data	Preço	Qtnd	Total (R\$)	Oper.	Lucro (R\$)	Data	Preço	Qtnd	Total (R\$)	Oper.	Lucro (R\$)
06/07/2010	1028,06	97	99721,83	Compra		22/10/2012	1433,82	173	248050,85	Compra	
07/07/2010	1060,27	97	102846,19	Vende	3124,37	23/10/2012	1413,11	173	244468,03	Vende	-3582,82
16/07/2010	1064,88	96	102228,48	Compra		24/10/2012	1408,75	173	243713,75	Compra	
19/07/2010	1071,25	96	102840,00	Vende	611,52	08/11/2012	1377,51	173	238309,23	Vende	-5404,52
20/07/2010	1083,48	95	102930,60	Compra		14/11/2012	1355,49	176	238566,24	Compra	
23/07/2010	1102,66	95	104752,70	Vende	1822,11	21/11/2012	1391,03	176	244821,29	Vende	6255,05
28/07/2010	1106,13	95	105082,35	Compra		04/12/2012	1407,05	174	244826,71	Compra	
29/07/2010	1101,53	95	104645,35	Vende	-437,00	10/12/2012	1418,55	174	246827,71	Vende	2001,00
30/07/2010	1101,60	95	104652,00	Compra		12/12/2012	1428,48	173	247127,04	Compra	
02/08/2010	1125,86	95	106956,70	Vende	2304,70	18/12/2012	1446,79	173	250294,68	Vende	3167,64
06/08/2010	1121,64	95	106555,80	Compra		24/12/2012	1426,66	175	249665,51	Compra	
09/08/2010	1127,79	95	107140,05	Vende	584,25	26/12/2012	1419,83	175	248470,24	Vende	-1195,26
12/08/2010	1083,61	99	107277,39	Compra		28/12/2012	1402,43	177	248230,12	Compra	
18/08/2010	1094,16	99	108321,84	Vende	1044,45	02/01/2013	1462,42	177	258848,35	Vende	10618,23
19/08/2010	1075,63	101	108638,63	Compra		03/01/2013	1459,37	178	259767,86	Compra	
27/08/2010	1064,59	101	107523,59	Vende	-1115,04	04/01/2013	1466,47	178	261031,65	Vende	1263,80
30/08/2010	1048,92	102	106989,84	Compra		08/01/2013	1457,15	179	260829,85	Compra	
01/09/2010	1080,29	102	110189,58	Vende	3199,74	09/01/2013	1461,02	179	261522,58	Vende	692,73
07/09/2010	1091,84	101	110275,84	Compra		11/01/2013	1472,05	177	260552,86	Compra	
15/09/2010	1125,07	101	113632,06	Vende	3356,23	18/01/2013	1485,98	177	263018,46	Vende	2465,60
17/09/2010	1125,59	101	113684,59	Compra		23/01/2013	1494,81	176	263086,57	Compra	
20/09/2010	1142,71	101	115413,71	Vende	1729,12	01/02/2013	1513,17	176	266317,93	Vende	3231,36
22/09/2010	1134,28	102	115696,56	Compra		05/02/2013	1511,29	177	267498,34	Compra	
27/09/2010	1142,16	102	116500,32	Vende	803,76	06/02/2013	1512,12	177	267645,24	Vende	146,90
28/09/2010	1147,70	101	115917,70	Compra		20/02/2013	1511,95	177	267615,14	Compra	
04/10/2010	1137,03	101	114840,03	Vende	-1077,66	01/03/2013	1518,20	177	268721,39	Vende	1106,25
14/10/2010	1173,81	98	115033,39	Compra		04/03/2013	1525,20	176	268435,19	Compra	
18/10/2010	1184,71	98	116101,58	Vende	1068,19	13/03/2013	1554,52	176	273595,52	Vende	5160,33
20/10/2010	1178,17	99	116638,83	Compra		18/03/2013	1552,10	176	273169,60	Compra	
21/10/2010	1180,26	99	116845,74	Vende	206,91	19/03/2013	1548,34	176	272507,83	Vende	-661,76
27/10/2010	1182,45	99	117062,55	Compra		25/03/2013	1551,69	176	273097,43	Compra	
02/11/2010	1193,57	99	118163,42	Vende	1100,88	26/03/2013	1563,77	176	275223,52	Vende	2126,09
09/11/2010	1213,40	97	117699,80	Compra		04/04/2013	1559,98	176	274556,48	Compra	
18/11/2010	1196,69	97	116078,92	Vende	-1620,88	05/04/2013	1553,28	176	273377,29	Vende	-1179,19

19/11/2010	1199,73	97	116373,81	Compra		09/04/2013	1568,61	174	272938,14	Compra	
26/11/2010	1189,40	97	115371,80	Vende	-1002,01	10/04/2013	1587,73	174	276265,02	Vende	3326,88
29/11/2010	1187,76	97	115212,72	Compra		11/04/2013	1593,37	174	277246,38	Compra	
01/12/2010	1206,07	97	116988,78	Vende	1776,06	24/04/2013	1578,79	174	274709,47	Vende	-2536,91
06/12/2010	1223,12	96	117419,52	Compra		29/04/2013	1593,61	172	274100,92	Compra	
20/12/2010	1247,08	96	119719,68	Vende	2300,16	03/05/2013	1614,42	172	277680,25	Vende	3579,33
28/12/2010	1258,51	95	119558,45	Compra		09/05/2013	1626,67	171	278160,58	Compra	
03/01/2011	1271,87	95	120827,65	Vende	1269,20	13/05/2013	1633,77	171	279374,67	Vende	1214,10
06/01/2011	1273,85	95	121015,75	Compra		15/05/2013	1658,78	168	278675,04	Compra	
12/01/2011	1285,96	95	122166,20	Vende	1150,45	17/05/2013	1667,47	168	280134,96	Vende	1459,91
20/01/2011	1280,26	95	121624,70	Compra		21/05/2013	1669,16	168	280418,89	Compra	
21/01/2011	1283,35	95	121918,25	Vende	293,55	29/05/2013	1648,36	168	276924,48	Vende	-3494,41
28/01/2011	1276,34	95	121252,30	Compra		31/05/2013	1630,74	170	277225,80	Compra	
04/02/2011	1310,87	95	124532,65	Vende	3280,35	14/06/2013	1626,73	170	276544,10	Vende	-681,70
07/02/2011	1319,05	95	125309,75	Compra		18/06/2013	1651,81	167	275852,28	Compra	
09/02/2011	1320,88	95	125483,60	Vende	173,85	21/06/2013	1592,43	167	265935,82	Vende	-9916,46
14/02/2011	1332,32	94	125238,07	Compra		24/06/2013	1573,09	169	265852,20	Compra	
16/02/2011	1336,32	94	125614,07	Vende	376,00	27/06/2013	1613,20	169	272630,79	Vende	6778,59
23/02/2011	1307,40	96	125510,40	Compra		08/07/2013	1640,46	167	273956,81	Compra	
03/03/2011	1330,97	96	127773,12	Vende	2262,71	12/07/2013	1680,19	167	280591,72	Vende	6634,91
18/03/2011	1279,21	100	127921,00	Compra		16/07/2013	1676,26	167	279935,42	Compra	
21/03/2011	1298,38	100	129838,00	Vende	1917,00	19/07/2013	1692,09	167	282579,02	Vende	2643,60
22/03/2011	1293,77	100	129377,00	Compra		25/07/2013	1690,25	167	282271,75	Compra	
25/03/2011	1313,80	100	131380,00	Vende	2003,00	01/08/2013	1706,87	167	285047,29	Vende	2775,54
28/03/2011	1310,19	101	132329,18	Compra		06/08/2013	1697,37	168	285158,16	Compra	
29/03/2011	1319,44	101	133263,43	Vende	934,25	07/08/2013	1690,91	168	284072,89	Vende	-1085,27
08/04/2011	1328,17	100	132817,00	Compra		13/08/2013	1694,16	168	284618,89	Compra	
13/04/2011	1314,41	100	131441,00	Vende	-1376,00	14/08/2013	1685,39	168	283145,52	Vende	-1473,36
14/04/2011	1314,52	100	131452,00	Compra		15/08/2013	1661,32	170	282424,39	Compra	
15/04/2011	1319,68	100	131968,01	Vende	516,00	22/08/2013	1656,96	170	281683,19	Vende	-741,20
18/04/2011	1305,14	101	131819,14	Compra		27/08/2013	1630,48	173	282073,04	Compra	
27/04/2011	1355,66	101	136921,66	Vende	5102,52	05/09/2013	1655,08	173	286328,83	Vende	4255,80
03/05/2011	1356,62	101	137018,62	Compra		06/09/2013	1655,17	173	286344,42	Compra	
11/05/2011	1342,08	101	135550,08	Vende	-1468,54	09/09/2013	1671,71	173	289205,82	Vende	2861,41
12/05/2011	1348,65	101	136213,65	Compra		24/09/2013	1697,42	170	288561,41	Compra	
16/05/2011	1329,47	101	134276,47	Vende	-1937,19	02/10/2013	1693,87	170	287957,90	Vende	-603,51
17/05/2011	1328,98	101	134226,98	Compra		08/10/2013	1655,45	174	288048,29	Compra	
18/05/2011	1340,68	101	135408,69	Vende	1181,71	18/10/2013	1744,50	174	303543,00	Vende	15494,71
19/05/2011	1343,60	100	134360,00	Compra		22/10/2013	1754,67	173	303557,92	Compra	
26/05/2011	1325,69	100	132568,99	Vende	-1791,00	28/10/2013	1762,11	173	304845,03	Vende	1287,11
27/05/2011	1331,10	100	133110,00	Compra		30/10/2013	1763,31	173	305052,64	Compra	
31/05/2011	1345,20	100	134520,00	Vende	1410,00	04/11/2013	1767,93	173	305851,90	Vende	799,26
07/06/2011	1284,94	105	134918,69	Compra		05/11/2013	1762,97	174	306756,77	Compra	
14/06/2011	1287,87	105	135226,35	Vende	307,66	14/11/2013	1790,62	174	311567,88	Vende	4811,10
16/06/2011	1267,64	106	134369,84	Compra		19/11/2013	1787,87	174	311089,38	Compra	
22/06/2011	1287,14	106	136436,84	Vende	2067,00	22/11/2013	1804,76	174	314028,24	Vende	2938,86
28/06/2011	1296,67	106	137447,02	Compra		03/12/2013	1795,15	175	314151,25	Compra	
01/07/2011	1339,67	106	142005,02	Vende	4558,00	16/12/2013	1786,54	175	312644,51	Vende	-1506,75
11/07/2011	1319,49	107	141185,43	Compra		19/12/2013	1809,60	173	313060,80	Compra	
25/07/2011	1337,43	107	143105,02	Vende	1919,59	26/12/2013	1842,02	173	318669,46	Vende	5608,67
15/08/2011	1204,49	119	143334,31	Compra		03/01/2014	1831,37	174	318658,38	Compra	
18/08/2011	1140,65	119	135737,35	Vende	-7596,96	23/01/2014	1828,46	174	318152,03	Vende	-506,35
19/08/2011	1123,53	121	135947,13	Compra		27/01/2014	1781,56	178	317117,69	Compra	
26/08/2011	1176,80	121	142392,81	Vende	6445,67	28/01/2014	1792,50	178	319065,00	Vende	1947,31
01/09/2011	1204,42	118	142121,57	Compra		03/02/2014	1741,89	183	318765,87	Compra	
08/09/2011	1185,90	118	139936,20	Vende	-2185,36	06/02/2014	1773,43	183	324537,70	Vende	5771,83

09/09/2011	1154,23	121	139661,83	Compra		12/02/2014	1819,26	179	325647,54	Compra	
19/09/2011	1204,09	121	145694,89	Vende	6033,06	13/02/2014	1829,83	179	327539,56	Vende	1892,02
22/09/2011	1129,56	129	145713,25	Compra		21/02/2014	1836,25	178	326852,50	Compra	
27/09/2011	1175,38	129	151624,02	Vende	5910,77	26/02/2014	1845,16	178	328438,49	Vende	1585,99
29/09/2011	1160,40	131	152012,40	Compra		03/03/2014	1845,73	178	328539,94	Compra	
10/10/2011	1194,89	131	156530,59	Vende	4518,19	04/03/2014	1873,91	178	333555,99	Vende	5016,05
12/10/2011	1207,25	130	156942,50	Compra		06/03/2014	1877,03	178	334111,35	Compra	
18/10/2011	1225,38	130	159299,40	Vende	2356,90	10/03/2014	1877,17	178	334136,27	Vende	24,92
20/10/2011	1215,39	131	159216,09	Compra		12/03/2014	1868,20	179	334407,79	Compra	
24/10/2011	1254,19	131	164298,88	Vende	5082,79	18/03/2014	1872,25	179	335132,75	Vende	724,96
25/10/2011	1229,05	133	163463,66	Compra		20/03/2014	1872,01	179	335089,79	Compra	
27/10/2011	1284,59	133	170850,47	Vende	7386,81	25/03/2014	1865,62	179	333945,98	Vende	-1143,81
04/11/2011	1253,23	137	171692,51	Compra		28/03/2014	1857,62	179	332513,98	Compra	
08/11/2011	1275,92	137	174801,05	Vende	3108,54	01/04/2014	1885,52	179	337508,08	Vende	4994,10
09/11/2011	1229,10	142	174532,20	Compra		08/04/2014	1851,96	183	338908,67	Compra	
15/11/2011	1257,81	142	178609,03	Vende	4076,83	15/04/2014	1842,98	183	337265,34	Vende	-1643,34
22/11/2011	1188,04	150	178206,01	Compra		16/04/2014	1862,31	181	337078,12	Compra	
28/11/2011	1192,55	150	178882,51	Vende	676,50	23/04/2014	1875,39	181	339445,59	Vende	2367,47
01/12/2011	1244,58	144	179219,51	Compra		29/04/2014	1878,33	180	338099,39	Compra	
05/12/2011	1257,08	144	181019,51	Vende	1800,00	30/04/2014	1883,95	180	339110,99	Vende	1011,60
08/12/2011	1234,35	147	181449,45	Compra		09/05/2014	1878,48	181	340004,88	Compra	
09/12/2011	1255,19	147	184512,92	Vende	3063,47	12/05/2014	1896,65	181	343293,65	Vende	3288,78
12/12/2011	1236,47	149	184234,03	Compra		15/05/2014	1870,85	183	342365,55	Compra	
21/12/2011	1243,72	149	185314,28	Vende	1080,25	23/05/2014	1900,53	183	347797,00	Vende	5431,45
30/12/2011	1257,60	147	184867,20	Compra		28/05/2014	1909,78	182	347579,97	Compra	
12/01/2012	1295,50	147	190438,50	Vende	5571,30	06/06/2014	1949,44	182	354798,07	Vende	7218,10
13/01/2012	1289,09	148	190785,31	Compra		11/06/2014	1943,89	183	355731,87	Compra	
20/01/2012	1315,38	148	194676,24	Vende	3890,93	12/06/2014	1930,11	183	353210,13	Vende	-2521,75
24/01/2012	1314,65	148	194568,20	Compra		13/06/2014	1936,16	182	352381,13	Compra	
25/01/2012	1326,06	148	196256,89	Vende	1688,69	19/06/2014	1959,48	182	356625,36	Vende	4244,23
26/01/2012	1318,43	149	196446,08	Compra		27/06/2014	1960,96	182	356894,71	Compra	
03/02/2012	1344,90	149	200390,10	Vende	3944,03	01/07/2014	1973,32	182	359144,23	Vende	2249,52
10/02/2012	1342,64	149	200053,36	Compra		08/07/2014	1963,71	183	359358,92	Compra	
16/02/2012	1358,04	149	202347,97	Vende	2294,60	09/07/2014	1972,83	183	361027,88	Vende	1668,96
17/02/2012	1361,23	149	202823,27	Compra		10/07/2014	1964,68	184	361501,13	Compra	
23/02/2012	1363,46	149	203155,53	Vende	332,27	16/07/2014	1981,57	184	364608,87	Vende	3107,74
05/03/2012	1364,33	149	203285,16	Compra		17/07/2014	1958,12	186	364210,32	Compra	
14/03/2012	1394,28	149	207747,72	Vende	4462,56	21/07/2014	1973,63	186	367095,18	Vende	2884,86
21/03/2012	1402,89	148	207627,72	Compra		22/07/2014	1983,53	185	366953,06	Compra	
26/03/2012	1416,51	148	209643,48	Vende	2015,76	23/07/2014	1987,01	185	367596,85	Vende	643,80
10/04/2012	1358,59	154	209222,85	Compra		25/07/2014	1978,34	186	367971,23	Compra	
20/04/2012	1378,53	154	212293,62	Vende	3070,77	28/07/2014	1978,91	186	368077,27	Vende	106,03
23/04/2012	1366,94	155	211875,69	Compra		31/07/2014	1930,67	191	368757,98	Compra	
26/04/2012	1399,98	155	216996,90	Vende	5121,21	04/08/2014	1938,99	191	370347,09	Vende	1589,11
27/04/2012	1403,36	155	217520,80	Compra		05/08/2014	1920,21	193	370600,52	Compra	
30/04/2012	1397,91	155	216676,06	Vende	-844,74	21/08/2014	1992,37	193	384527,41	Vende	13926,89
01/05/2012	1405,82	154	216496,27	Compra		26/08/2014	2000,02	192	384003,84	Compra	
04/05/2012	1369,10	154	210841,40	Vende	-5654,88	29/08/2014	2003,37	192	384647,04	Vende	643,20
01/06/2012	1278,04	165	210876,61	Compra		04/09/2014	1997,65	192	383548,80	Compra	
06/06/2012	1315,13	165	216996,45	Vende	6119,84	09/09/2014	1988,44	192	381780,47	Vende	-1768,34
08/06/2012	1325,66	164	217408,25	Compra		12/09/2014	1985,54	193	383209,23	Compra	
19/06/2012	1357,98	164	222708,72	Vende	5300,47	17/09/2014	2001,57	193	386303,00	Vende	3093,77
21/06/2012	1325,51	168	222685,68	Compra		29/09/2014	1977,80	195	385671,01	Compra	
22/06/2012	1335,02	168	224283,36	Vende	1597,68	06/10/2014	1964,82	195	383139,89	Vende	-2531,12
26/06/2012	1319,99	170	224398,30	Compra		21/10/2014	1941,28	197	382432,17	Compra	
29/06/2012	1362,16	170	231567,21	Vende	7168,91	30/10/2014	1994,65	197	392946,05	Vende	10513,89

11/07/2012	1341,45	172	230729,39	Compra		06/11/2014	2031,21	194	394054,73	Compra	
20/07/2012	1362,66	172	234377,53	Vende	3648,13	10/11/2014	2038,26	194	395422,44	Vende	1367,71
24/07/2012	1338,31	175	234204,26	Compra		11/11/2014	2039,68	194	395697,93	Compra	
27/07/2012	1385,97	175	242544,74	Vende	8340,48	17/11/2014	2041,32	194	396016,07	Vende	318,14
07/08/2012	1401,35	173	242433,55	Compra		18/11/2014	2051,80	193	395997,41	Compra	
10/08/2012	1405,87	173	243215,51	Vende	781,96	20/11/2014	2052,75	193	396180,75	Vende	183,34
16/08/2012	1415,51	172	243467,72	Compra		21/11/2014	2063,50	192	396192,00	Compra	
17/08/2012	1418,16	172	243923,53	Vende	455,80	24/11/2014	2069,41	192	397326,70	Vende	1134,70
27/08/2012	1410,44	173	244006,11	Compra		25/11/2014	2067,03	192	396869,77	Compra	
04/09/2012	1404,94	173	243054,61	Vende	-951,50	26/11/2014	2072,83	192	397983,37	Vende	1113,61
06/09/2012	1432,12	170	243460,40	Compra		02/12/2014	2066,55	192	396777,61	Compra	
17/09/2012	1461,19	170	248402,29	Vende	4941,89	08/12/2014	2060,31	192	395579,53	Vende	-1198,08
19/09/2012	1461,05	170	248378,51	Compra		12/12/2014	2002,33	198	396461,33	Compra	
24/09/2012	1456,89	170	247671,30	Vende	-707,21	22/12/2014	2078,54	198	411550,93	Vende	15089,60
25/09/2012	1441,59	172	247953,47	Compra		24/12/2014	2081,88	198	412212,22	Compra	
01/10/2012	1444,49	172	248452,28	Vende	498,80	26/12/2014	2088,77	198	413576,46	Vende	1364,25
04/10/2012	1461,40	170	248438,00	Compra		30/12/2014	2080,35	198	411909,32	Compra	
18/10/2012	1457,34	170	247747,79	Vende	-690,21	31/12/2014	2058,90	198	407662,18	Vende	-4247,14