

Alexandre Pimenta

Métodos Automatizados para Investimento no Mercado de Ações via Inteligência Computacional

Tese de Doutorado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia Elétrica da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Doutor em Engenharia Elétrica.

Universidade Federal de Minas Gerais – UFMG

Programa de Pós-Graduação em Engenharia Elétrica – PPGEE

Orientador: Prof. Dr. Eduardo Gontijo Carrano

Coorientador: Prof. Dr. Frederico Gadelha Guimarães

Belo Horizonte–MG

25 de setembro de 2017

Dedicatória

Este trabalho é dedicado a todas as instituições federais de ensino do Brasil.

Agradecimentos

A princípio, gostaria de agradecer ao Professor Doutor Eduardo Gontijo Carrano pelo seu compromisso com a educação e com a qualidade da orientação. Meu orientador esteve presente a todo momento, desde o início, ajudando de forma decisiva a superar cada etapa desta difícil jornada. Não há texto científico que consiga traduzir a minha gratidão.

Ao Professor Doutor Frederico Gadelha Guimarães, que, apesar dos milhares de compromissos e da mesa cheia de trabalho, sempre esteve bem humorado e disposto a colaborar, a contribuir de forma muito significativa para o desenvolvimento do projeto de doutorado. Está guardada a sua mensagem via *Whatsapp*: “Viu a boa notícia?”. Estávamos aguardando há 7 meses o retorno da avaliação do nosso artigo. Por todo esse tempo dedicado, fica a minha gratidão.

Aos brilhantes professores da Engenharia Elétrica da Universidade Federal de Minas Gerais (UFMG), com os quais tive a oportunidade de aprender, dentre eles, Antônio Braga, João Vasconcelos e Felipe Campelo.

À minha esposa, Marianna Costa Mattos, e, à minha filha, Maria Costa Pimenta, que acompanharam o outro lado do doutorado, aquele que só quem convive na mesma casa sabe como é. Agradeço pela paciência, pelas palavras de harmonia e pela compreensão. Vocês duas foram o pilar emocional dessa jornada.

Aos meus pais, Maria Aparecida Ferreira Pimenta e Milton Linhares Pimenta. Toda caminhada tem um começo, e tudo começou com vocês. Sou grato por toda dedicação e amor recebidos de vocês.

Aos meus irmãos, Adriana Pimenta, Alessandra Pimenta e Dênis Eduardo Pimenta, pela paciência que tiveram comigo nesse período e por entenderem minhas ausências.

Aos apoios nas traduções e revisões de texto, feitos por Erika, Gregório, Thaís, Camila, Cíntia e Rosklin, que tiveram paciência e dedicação.

Ao amigo Ciniro, companheiro dessa jornada acadêmica, o qual sempre contribuiu com o debate de ideias, com o seu brilhantismo e independência intelectual. Em momentos de dificuldade sempre esteve presente.

Ao amigo Washington, professor do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG), pessoa íntegra e principal incentivador do meu doutorado. Foram muitas as vezes que recorri aos seus conhecimentos para resolver questões difíceis. Foram muitas conversas e muitos debates sobre computação e economia.

Ao amigo Robson, professor do IFMG, pessoa dedicada à educação, um incentivador da

qualificação docente no IFMG.

Ao amigo Neimar, incansável batalhador da pesquisa e pós-graduação no IFMG. Nossa Instituição está melhor com sua dedicação e trabalho.

Ao amigo Carlos Bernardes, amigo de longa data e competente gestor do IFMG.

Em nome do coordenador, Bruno Ferreira, gostaria também de fazer um agradecimento especial a todo corpo docente do curso de Ciência da Computação do IFMG *Campus* Formiga, o qual compreende a importância da qualificação docente e incetiva seus professores a participarem de programas de qualidade.

Ao Instituto Federal de Minas Gerais, onde sou professor, instituição tão nova e tão promissora. Da sua maneira vem mostrando, a força que uma educação de qualidade desempenha “nos interiores dessas Minas Gerais”. Sou grato por participar deste time.

Aos amigos, que de certa maneira acompanharam este trabalho e tiveram paciência em escutar as minhas ideias: Afonso, Álvaro, Alisson, Adriano, André, Plácido, Giuliano, Aldo, Paulo Lima, Hugo Bravo, Rosklin, Charles, Mardem Junior, Romero, Adriano, José Aparecido (Zezé), Dênis Quaresma, Aguinaldo e César.

E, por último, quero agradecer a oportunidade que a vida me deu, pois fui estudante de escola pública desde o primário e tive sorte de, nessa jornada, encontrar com excelentes professores. Passei pela Universidade Federal de Uberlândia (UFU), pela Universidade Federal do Rio Grande do Sul (UFRGS) e, agora, conheci de perto a UFMG. Para mim, um grande sonho realizado.

Epígrafe

“O impulso para a aquisição, a busca do ganho, do dinheiro, da maior quantidade possível de dinheiro, não tem nada a ver com o capitalismo. Tal impulso existe e tem existido entre garçons, médicos, artistas, cocheiros, prostitutas, funcionários desonestos, soldados, nobres, cruzados, apostadores e mendigos. Pode-se dizer que ele tem sido comum a todos os tipos de homens, de todas condições, em todas as épocas e países do planeta, onde quer que a possibilidade objetiva dele exista ou tenha existido.[...] A auri sacra fames é tão antiga quanto a história do homem”
(WEBER, 1930)

“Três quartos das demandas existentes no mundo são românticos; baseadas em visões, idealismos, esperanças e afeições; e a regulação da bolsa é, em essência, a regulação da imaginação e do coração” (RUSKIN, 1860)

Sumário

	Sumário	6
	Lista de ilustrações	11
	Lista de tabelas	14
1	INTRODUÇÃO	1
1.1	Contextualização	1
1.1.1	A computação e o mercado financeiro	1
1.1.2	Ações e Bolsa de Valores	2
1.1.3	Prazo Operacional	3
1.2	Desafios	4
1.3	Justificativa	5
1.4	Objetivos	6
1.5	Organização do Trabalho	6
2	REVISÃO BIBLIOGRÁFICA	7
2.1	Métodos do Mercado	7
2.1.1	Teoria de Dow	7
2.1.2	Análise Técnica	9
2.1.3	Análise Fundamentalista	11
2.2	Métodos Convencionais	11
2.3	Sistemas computacionais que fazem uso de inteligência computacional	12
2.3.1	Sistemas Inteligentes para Previsão de Séries Financeiras (SIPSF)	12
2.3.2	A Programação Genética e a Previsão de Séries Financeiras	13
3	FERRAMENTAS	17
3.1	Os Indicadores Numéricos da Análise Técnica	17
3.1.1	Média Móvel Simples (MMS)	18
3.1.2	Média Móvel Exponencial (MME)	20
3.1.3	Acumulação e Distribuição (AD)	21
3.1.4	Oscilador de Chaikin (OC)	22
3.1.5	Volatilidade do Oscilador de Chaikin (VC)	23
3.1.6	Williams %R (WR)	24
3.1.7	Volume Taxa de Variação (VROC)	25
3.1.8	Preço Taxa de Variação (PROC)	26

3.1.9	Bandas de Bollinger (BB)	27
3.1.10	Saldo de Volume (OBV)	28
3.1.11	Índice de Força Relativa (IFR)	29
3.1.12	Estocástico (E)	30
3.2	Deteccção e Eliminação de <i>Outliers</i>	32
3.2.1	Método do Quartil	33
3.3	Seleção de Características	34
3.3.1	Redução da Dimensionalidade	34
3.3.2	Classes de métodos de seleção de características	34
3.4	Programação Genética	36
3.4.1	Representação dos Programas	36
3.4.2	Geração da População Inicial	37
3.4.3	Avaliação dos Indivíduos	38
3.4.4	Métodos de Seleção	38
3.4.5	Operadores Genéticos	38
3.5	Algoritmos Evolucionários Multiobjetivo	39
3.5.1	NSGA	41
3.5.2	NSGA-II	43
3.6	Comitê de Decisão	47
4	SISTEMAS PROPOSTOS	50
4.1	Considerações Iniciais	50
4.2	Sistema Automatizado de Investimento via Seleção Multiobjetivo de Regras (SisSM)	51
4.2.1	Visão Geral do Sistema do SisSM	52
4.2.2	Algoritmo geral do SisSM	54
4.2.3	Seleção de Regras Fase I do SisSM	56
4.2.4	Seleção de Regras Fase II do SisSM	60
4.2.5	Funcionamento e Formação do Comitê de Decisão do SisSM	60
4.2.6	Sistema de negociação do SisSM	62
4.2.7	Exemplo de funcionamento do módulo de negociação do SisSM	66
4.2.8	Exemplo de funcionamento do SisSM	68
4.3	Sistema Automatizado de Investimento via Programação Genética Multiobjetivo	79
4.3.1	Visão Geral do Sistema do Sistema Automatizado de Investimento via Programação Genética Multiobjetivo (SisPG)	80
4.3.2	Algoritmo geral do SisPG	82
4.3.3	Deteccção e Remoção de <i>Outliers</i> do SisPG	83
4.3.4	Seleção de Características do SisPG	85
4.3.5	Programação Genética do SisPG	88

4.3.6	Comitê de Decisão do SisPG	91
4.3.7	Sistema de negociação do SisPG	91
5	RESULTADOS	92
5.1	Estratégias de Testes do SisSM	93
5.2	Resultados do SisSM	93
5.2.1	Teste de Estresse no SisSM	102
5.3	Estratégias de Testes do SisPG	102
5.4	Resultados do SisPG	103
5.4.1	Teste de Estresse no SisPG	111
5.5	Comparação entre os Resultados do SisPG e SisSM	111
6	CONCLUSÕES E PROPOSTAS DE CONTINUIDADE	113
6.1	Conclusões	113
6.1.1	Sistema Automatizado de Investimento via Seleção Multiobjetivo de Regras (SisSM)	113
6.1.2	Sistema Automatizado de Investimento via Programação Genética Multiobjetivo (SisPG)	113
6.1.3	Adaptação a outros mercados de capitais	113
6.1.4	Módulo de Negociação	114
6.1.5	Regras Lógicas e Indicadores da Análise Técnica	114
6.1.6	Estudos sobre a Bovespa	114
6.2	Proposta de Continuidade	114
6.3	Produção Bibliográfica Durante o Doutorado	114
	REFERÊNCIAS	116

Resumo

Obter ganhos financeiros na Bolsa de Valores por meio de estratégias automatizadas é uma área de forte interesse da academia e do mercado. Várias técnicas têm sido empregadas para tratar esse problema, como preditores da família ARCH e GARCH, redes neurais, lógica *fuzzy*, programação genética e métodos convencionais de mercado. Esse trabalho apresenta dois sistemas computacionais desenvolvidos para essa finalidade. O primeiro sistema se baseia na seleção ótima de regras de compra e de venda de ações. Já o segundo combina métodos convencionais, programação genética e otimização multiobjetivo. Os dois sistemas propostos têm a finalidade de obter lucro no mercado financeiro, identificando de forma automática os melhores momentos de compra e de venda de ações em uma janela de tempo determinada. Diferentemente de outros tipos de sistemas de investimento, nos quais o objetivo é tentar prever com exatidão o preço da ação no futuro, os sistemas desenvolvidos nesse trabalho buscam identificar regiões prolíficas para compra e venda de ações. Os sistemas foram testados em seis séries temporais históricas de ativos representativos da Bolsa de Valores do Brasil: a BOVESPA. Esses ativos estão distribuídos em quatro segmentos distintos da economia brasileira. Os sistemas apresentaram ganhos superiores à valorização dos ativos testados no período.

Palavras-chaves: Computação Evolucionária. Programação Genética. Otimização Multiobjetivo. Análise Técnica. Mercado de Ações.

Abstract

Automated investment in stock market is an area of strong interest of the academy and professional investors. In addition to conventional market methods, various sophisticated techniques have been employed to deal with the problem, such as ARCH/GARCH predictors, artificial neural networks, and fuzzy logic. This work presents two computational systems developed for this purpose. The first system is based on the optimal selection of rules of purchase and sale of shares. The second one combines conventional methods, genetic programming and multiobjective optimization. Both systems aim to reach profit in the financial market, automatically identifying the best moments for purchasing and selling shares in a given time window. Unlike other types of investment systems, in which the objective is to try to predict the exact price of the stock in the future, the systems developed in this work seek to identify suitable regions for purchase or sale of shares. The systems were tested in six historical time series of assets representative of the Brazilian stock market: BOVESPA. These assets are distributed in four distinct segments of the Brazilian economy. The systems presented higher gains than the valuation of the assets tested in the period.

Key-words: Evolutionary Computation. Genetic Programming. MultiObjective Optimization. Technical Analysis. Stock market.

Lista de ilustrações

Figura 1 – Componente de um gráfico de <i>Candles</i>	10
Figura 2 – Exemplo de gráfico de <i>Candles</i> aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 30/09/2016 a 01/12/2016. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	10
Figura 3 – Combinação de técnicas de sistemas computacionais financeiros	13
Figura 4 – Exemplo de MMS aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	20
Figura 5 – Exemplo de MME aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	21
Figura 6 – Exemplo de AD aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	22
Figura 7 – Exemplo de OC aplicado em uma série financeira	23
Figura 8 – Exemplo de VC aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	24
Figura 9 – Exemplo de WR aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	25
Figura 10 – Exemplo de VROC aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	26
Figura 11 – Exemplo de PROC aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	27

Figura 12 – Exemplo de BB aplicadas na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos(XP, 2017).	29
Figura 13 – Exemplo de OBV aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	30
Figura 14 – Exemplo de IFR aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	31
Figura 15 – Exemplo de Estocástico (E) aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)	32
Figura 16 – Técnicas de Redução de Dimensionalidade	35
Figura 17 – <i>Overfitting</i> em função da dimensionalidade	35
Figura 18 – Representação de um indivíduo pela estrutura árvore	37
Figura 19 – Cruzamento por ponto de corte	39
Figura 20 – Mutação aleatória	40
Figura 21 – Exemplo de ordenação pelo conceito de dominância Pareto	41
Figura 22 – Princípio de Seleção do NSGA-II	45
Figura 23 – Fases do SisSM	52
Figura 24 – Esquema Geral do SisSM. Neste esquema: $D[\{t_I, \dots, t_F\}]$ é a série financeira inicial t_I e final t_F ; $\Delta_F > 0$, $\Delta_T > 0$, e $\Delta_E > 0$ tamanho das janelas de seleção de características, treinamento e negociação, respectivamente	53
Figura 25 – Diagrama de transição do módulo de negociação. Na figura, Fora (início) e Fora (fim) representam o estado (inicial) e o estado (final) do diagrama . . .	64
Figura 26 – BOVA11, exemplo do sistema de negociação	66
Figura 27 – Exemplo de venda a descoberto, Ação - CMIG4	77
Figura 28 – Fases do SisPG	79
Figura 29 – Esquema Geral do SisPG. Neste esquema: $D[\{t_I, \dots, t_F\}]$ é a série financeira inicial t_I e final t_F ; $\Delta_F > 0$, $\Delta_T > 0$, e $\Delta_E > 0$ tamanho das janelas de Seleção de Características, Treinamento e Negociação, respectivamente	81
Figura 30 – Exemplo de dados (BBAS3 entre 7 de maio de 2013 e 19 de junho de 2013). Antes (linha contínua preta) e Depois (linha tracejada vermelha) de aplicar o algoritmo de detecção e remoção de <i>Outlier</i>	84
Figura 31 – Exemplo de indivíduo	89

Figura 32 – BBAS3 - Valor da Ação X Desempenho do SisSM	96
Figura 33 – BOVA11 - Valor da Ação X Desempenho do SisSM	97
Figura 34 – CMIG4 - Valor da Ação X Desempenho do SisSM	98
Figura 35 – GGBR4 - Valor da Ação X Desempenho do SisSM	99
Figura 36 – EMBR3- Valor da Ação X Desempenho do SisSM	100
Figura 37 – VALE5 - Valor da Ação X Desempenho do SisSM	101
Figura 38 – BBAS3 - Valor da Ação X Desempenho do SisPG	105
Figura 39 – BOVA11 - Valor da Ação X Desempenho do SisPG	106
Figura 40 – CMIG4 - Valor da Ação X Desempenho do SisPG	107
Figura 41 – GGBR4 - Valor da Ação X Desempenho do SisPG	108
Figura 42 – EMBR3- Valor da Ação X Desempenho do SisPG	109
Figura 43 – VALE5 - Valor da Ação X Desempenho do SisPG	110

Lista de tabelas

Tabela 1 – Principais indicadores técnicos	19
Tabela 2 – Conjunto de números inteiros ordenados	33
Tabela 3 – Indicadores Numéricos da Análise Técnica (AT)	57
Tabela 4 – Regras lógicas de compra baseadas em indicadores numéricos da AT do SisSM	58
Tabela 5 – Regras lógicas de venda baseadas em indicadores numéricos da AT do SisSM	59
Tabela 6 – Exemplo do Sistema de Negociação com o BOVA11	67
Tabela 7 – Parâmetros adotados no exemplo de funcionamento do SisSM	68
Tabela 8 – Regras lógicas baseadas em indicadores numéricos da AT selecionadas na fase II do SisSM	69
Tabela 9 – Exemplo de funcionamento do SisSM, evolução do capital, iteração I	72
Tabela 10 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração I	72
Tabela 11 – Exemplo de funcionamento do SisSM, evolução do capital, iteração II . . .	72
Tabela 12 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração II	73
Tabela 13 – Exemplo de funcionamento do SisSM, evolução do capital, iteração III . . .	73
Tabela 14 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração III	73
Tabela 15 – Exemplo de funcionamento do SisSM, evolução do capital, iteração IV . . .	74
Tabela 16 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração IV	74
Tabela 17 – Exemplo de funcionamento do SisSM, evolução do capital, iteração V . . .	75
Tabela 18 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração V	75
Tabela 19 – Exemplo de funcionamento do SisSM, evolução do capital, iteração VI . . .	76
Tabela 20 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração VI	76
Tabela 21 – Exemplo de funcionamento do SisSM, evolução do capital, iteração VII . .	77
Tabela 22 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração VII	78
Tabela 23 – Detalhamento da terceira iteração, exemplo de venda a descoberto, Ação - CMIG4	79
Tabela 24 – Regras lógicas de compra baseadas em indicadores numéricos da AT do SisPG	86
Tabela 25 – Regras lógicas de venda baseadas em indicadores numéricos da AT do SisPG	87
Tabela 26 – Tomada de decisão individual	89
Tabela 27 – Configuração da Programação Genética (PG) Multiobjetivo	91

Tabela 28 – Código e descrição das ações utilizadas nos testes do SisSM e SisPG	92
Tabela 29 – Parâmetros adotados nos testes do SisSM	93
Tabela 30 – Retorno financeiro obtido pelo SisSM e as duas estratégias apresentadas . .	94
Tabela 31 – Retorno financeiro obtido pelo par de regras da AT utilizadas no SisSM . . .	95
Tabela 32 – Retorno financeiro obtido pelo SisSM e as duas estratégias apresentadas (Período de crise)	102
Tabela 33 – Parâmetros adotados nos testes do SisPG	103
Tabela 34 – Retorno financeiro obtido pelo SisPG e as três estratégias apresentadas . . .	104
Tabela 35 – Retorno financeiro obtido pelo SisPG e as três estratégias apresentadas (Pe- ríodo de Crise)	111
Tabela 36 – Comparação entre os retornos financeiros obtidos pelo SisPG e os retornos financeiros obtidos pelo SisSM no primeiro experimento	112
Tabela 37 – Comparação entre os retornos financeiros obtidos pelo SisPG e os retornos financeiros obtidos pelo SisSM no segundo experimento	112

Lista de Algoritmos

1	Pseudocódigo do algoritmo PG	36
2	Pseudocódigo do algoritmo <i>Fitness</i> -NSGA:	43
3	Pseudocódigo do algoritmo NSGA-II	44
4	Pseudocódigo do <i>Fast Non-Dominated Sorting</i> (<i>P</i>)	46
5	Pseudocódigo do <i>Crowding-distance-assignment</i> (<i>P</i>)	46
6	Pseudocódigo do SisSM	55
7	Pseudocódigo da Seleção de Regras Fase I	57
8	Pseudocódigo da Seleção de Regras Fase II	60
9	Pseudocódigo do Comitê de Decisão	62
10	Pseudocódigo do Sistema de Negociação	65
11	Pseudocódigo do SisPG	82
12	Pseudocódigo do Remove <i>Outlier</i>	84
13	Pseudocódigo da Seleção de Características	88
14	Pseudocódigo do Treinamento	90

Lista de abreviaturas e siglas

AD Acumulação e Distribuição

AEs Algoritmos Evolucionários

AEM Algoritmos Evolucionários Multiobjetivo

AF Análise Fundamentalista

AG Algoritmo Genético

AM Aprendizagem de Máquina

ARCH *Autoregressive Conditional Heteroskedasticity*

ARFIMA *Autoregressive Fractionally Integrated Moving Average*

ARIMA *Autoregressive Integrated Moving Average*

ARMA *Autoregressive Moving Average Models*

AT Análise Técnica

BB Bandas de Bollinger

CE Computação Evolutiva

E Estocástico

GARCH *Generalized AutoRegressive Conditional Heteroskedasticity*

HME Hipótese do Mercado Eficiente

HFT *High Frequency Trading*

IFR Índice de Força Relativa

IC Inteligência Computacional

MMS Média Móvel Simples

MME Média Móvel Exponencial

OC Oscilador de Chaikin

OBV Saldo de Volume

PE Pregão Eletrônico

PROC Preço Taxa de Variação

PO Problema de Otimização Mono-Objetivo

POMO Problema de Otimização Multiobjetivo

PG Programação Genética

PVV Pregão Viva Voz

PSO *Particle Swarm Optimization*

SARMA Seasonal Autoregressive Moving Average

SARIMA Seasonal Autoregressive Integrated Moving Average

SisPG Sistema Automatizado de Investimento via Programação Genética Multiobjetivo

SisSM Sistema Automatizado de Investimento via Seleção Multiobjetivo de Regras

SIPSF Sistemas Inteligentes para Previsão de Séries Financeiras

SVM *Support Vector Machines*

VC Volatilidade do Oscilador de Chaikin

VROC Volume Taxa de Variação

WR Williams %R

1 Introdução

Este capítulo apresenta a contextualização desta pesquisa, bem como os motivos que justificam a tese e as soluções propostas. Também são apresentados os objetivos e a estrutura dos capítulos seguintes.

1.1 Contextualização

1.1.1 A computação e o mercado financeiro

Em 1965, Moore previu que o poder computacional iria dobrar a cada 18 meses ao longo das próximas décadas (SCHALLER, 1997). Essa teoria ficou conhecida como “a Lei de Moore” e teve impacto em praticamente todas as atividades humanas. No mercado financeiro não foi diferente, uma vez que houve uma confluência do surgimento de modelos econômicos com o crescimento do poder computacional. No trabalho de Kirilenko e Lo (2013), os autores apontam um paralelo entre o desenvolvimento de várias teorias em finanças com o desenvolvimento da computação. Com o objetivo de situar a proposta deste trabalho, é apresentada a seguir parte da categorização proposta por Kirilenko e Lo (2013) e é sugerida uma nova categoria: os Sistemas Inteligentes para Previsão de Séries Financeiras (SIPSF). Em comum, todas estas classes estão diretamente relacionadas com os avanços dos mercados de capitais e da computação:

- **Finanças Quantitativas:** as finanças quantitativas fazem uso intensivo da estatística e da econometria para modelar o mercado financeiro, tendo como um dos precursores Markowitz (1952), que foi o primeiro a propor uma negociação algorítmica para maximizar uma alocação de recursos de um investidor em uma carteira de investimento.
- **Fundos Indexados:** os fundos indexados são derivados das teorias desenvolvidas nas finanças quantitativas. Estes permitem que os investidores façam parte de um amplo mercado, diversificando seus investimentos, sem o custo de comprar individualmente cada ativo. Os modelos matemáticos desenvolvidos juntamente com o incremento do poder computacional fizeram com que esse tipo de investimento ampliasse consideravelmente.
- **Negociação de Arbitragem:** com o desenvolvimento da comunicação de dados junto à integração global dos mercados, foi possível construir algoritmos que vasculham imensas quantidades de preços no mercado em busca de precificação inadequada de ativos. “As oportunidades de arbitragem provêm de anomalias do mercado, onde a precificação de dois ativos idênticos contradiz a Lei do Preço Único – mesmo preço de mercado para ativos efetivamente idênticos. Algumas dessas anomalias são provocadas, principalmente, devido

à falta de liquidez, assimetria de informação, irracionalidade das expectativas, grandes volatilidades do mercado ou tempo de execução das ordens"(CHIARA et al., 2012).

- **Formação de Mercado, (*Market Making*)**: os formadores de mercado são operadores da Bolsa de Valores, a qual garante níveis adequados de liquidez e profundidade de ofertas, com impacto direto no processo de formação do preço. Existem contratos específicos entre empresas e *Market Makers* com o objetivo de prover liquidez em ativos com baixa liquidez.
- **Negociação em Alta Frequência, (*High Frequency Trading (HFT)*)**: recentemente, segundo Fabozzi, Focardi e Jonas (2011), uma outra forma de operar na bolsa foi introduzida: as HFT são negociações em alta frequência realizadas exclusivamente por computadores e algoritmos. As estratégias HFT são caracterizadas por um grande volume de negócios e pequenas margens de lucro por transação (ALDRIDGE, 2013). Essa estratégia aproveita ao máximo as inovações em computação e telecomunicações para negociar milhões de ações por dia.
- **SIPSF**: nesta categoria se enquadram todos os sistemas computacionais que fazem uso de técnicas da Inteligência Computacional (IC), com o objetivo de prever o preço de um ativo financeiro ou prever uma região de compra ou de venda em uma data futura. Esses sistemas fazem uso de algumas das seguintes técnicas: redes neurais, lógica *fuzzy*, computação evolucionária e algoritmos de otimização mono e multiobjetivo. Sua finalidade maior é obter retorno financeiro comprando um ativo com o preço baixo e vendendo com o preço alto, ou vice-versa. Esses sistemas identificam oportunidades na flutuação dos valores das ações no mercado financeiro.

O trabalho proposto nesta tese se enquadra na categoria dos **SIPSF**. Foram construídos dois SIPSF que utilizam técnicas de IC para atuar na compra e na venda de ações com o objetivo de obter retorno no mercado financeiro.

1.1.2 Ações e Bolsa de Valores

Ações são títulos de renda variável que representam frações de capital de uma determinada empresa. Uma **ação** é um título nominativo negociável que representa parte do capital social de uma determinada empresa, e seu preço varia de acordo com a *Lei da Oferta e da Procura*¹. O investimento em ações é, em geral, considerado uma oportunidade para formação de patrimônio e, ao mesmo tempo, financia o crescimento das empresas.

¹ Em economia, a Lei da Oferta e Procura é um modelo de determinação de preços num mercado. Num mercado em concorrência perfeita, o modelo argumenta que os agentes econômicos tomam decisões que variam o preço até que este seja tal que a quantidade procurada seja igual à quantidade oferecida, resultando daí um equilíbrio econômico em que não há incentivos para a alteração de quantidades ou preços (WIKIPEDIA, 2017b).

As Bolsas de Valores são os centros de negociações das ações e garantem a operacionalização das negociações. Estas são organizadas como associações com funções de interesse público. Portanto, a Bolsa de Valores é o local onde as ações das empresas são compradas e vendidas. No Brasil, este local é a Bolsa de Valores, Mercadorias e Futuros de São Paulo, conhecida como a BM&FBovespa.

As operações de compra e venda de ações na BMF&Bovespa evoluíram de Pregão Viva Voz (PVV), onde os ativos eram negociados verbalmente por operadores humanos, para o conceito de Pregão Eletrônico (PE), onde os ativos são negociados a distância sem intermediário humano, tendo apenas um sistema eletrônico como ferramenta de negociação. Para detalhes de funcionamento do PE e as tecnologias utilizadas recomenda-se consultar o site oficial BMF&Bovespa, na seção regulamentos e manuais (BMFBOVESPA, 2017).

1.1.3 Prazo Operacional

Pode-se classificar o investidor de acordo com a frequência de transações de compra e venda de ações (MURPHY, 1999; ELDER, 1993). Essa classificação determina o perfil do investidor.

1. **Buy and Hold** - São operações em que o prazo de retorno está previsto no período de anos;
2. **PositionTrader** - As operações de compra e a venda de ações se dão no prazo de semanas e meses;
3. **SwingTrader** - São operações que acontecem no prazo de 2 a 6 dias;
4. **DayTrader** - As operações de compra e venda de ações acontecem no mesmo dia. O investidor que realiza esse tipo de operação espera pequenas variações no preço em curto prazo, aproveitando apenas uma parte do movimento e realizando diversas operação ao longo do dia.

Todas as formas de operar na Bolsa de Valores apresentadas acima têm uma característica em comum: podem ser feitas por pessoas através do PE, ou seja, as ordens de compra e de venda podem ser disparadas por seres humanos. Recentemente, uma outra forma de operar na bolsa foi introduzida, as HFT, que são negociações realizadas por computadores e algoritmos (ALDRIDGE, 2013) e visam obter lucro através de um grande volume de negócios com uma pequena margem de rentabilidade em cada transação. Nesse caso, as transações são realizadas completamente por computadores, sem qualquer iteração humana.

1.2 Desafios

Fazer previsões em mercados financeiros, identificando seu comportamento futuro, não é uma tarefa simples, pois estes são constantemente influenciados por fatores econômicos, políticos ou sociais. No século XVIII, os japoneses desenvolveram o primeiro método de previsão dos preços de contratos futuros de arroz, chamado de *Candlestick* (MURPHY, 1999; ELDER, 1993). Desde este marco inicial, muitas técnicas vêm sendo desenvolvidas, combinando diversas áreas do conhecimento, como economia, matemática, estatística, física, psicologia e computação.

As teorias tradicionais de finanças, em quase sua totalidade, acreditam na racionalidade dos agentes econômicos e foram construídas sobre a validade da Hipótese do Mercado Eficiente (HME) (FAMA, 1970). Esta teoria pressupõe que os preços dos ativos refletem todas as informações disponíveis no mercado e, dado que todos os agentes possuem acesso às mesmas informações e todos possuem o mesmo propósito de lucrar, logo o valor de uma cotação será sempre justo. Segundo Fama (1970), esta teoria parte das seguintes premissas:

- os investidores são considerados indivíduos racionais que buscam o lucro a partir de informações públicas disponíveis;
- os preços são justos;
- o mercado tende a permanecer em equilíbrio, na medida em que sempre existirão vendedores e compradores;
- toda nova informação impacta diretamente de forma igual ao mercado e seus participantes;
- os preços atuais são independentes dos preços de ontem, pois toda informação disponível já foi incorporada a estes preços.

O matemático francês Bachelier (1900) foi o primeiro a realizar estudos sobre a aleatoriedade de preços de ativos financeiros. Ele identifica que o valor das ações tende a se comportar de forma aleatória, pois:

“...os fatores que influenciam as atividades nas bolsas são inúmeros e os eventos atuais ou esperados não apresentam relação aparente com a variação dos preços. Além das causas naturais das oscilações, também existem causas artificiais: a bolsa reage a si mesma e as negociações em curso são funções não só das negociações anteriores, mas também de seu relacionamento com o resto do mercado. A determinação dessa atividade depende de um número infinito de fatores: assim, é impossível contar com previsões matemáticas (BACHELIER, 1900).”

Bachelier (1900) institui o conceito de passeio aleatório nos mercados financeiros e seus estudos deram suporte à Teoria Moderna Financeira, que tem como pressuposto a HME. Sob

esta hipótese, a possibilidade de obtenção de lucros extraordinários é remota, pois a suposição de que um investidor com a posse de novas informações pudesse transformá-la em lucros seria reduzida. Para Samuelson (1965), o mercado eficiente é aquele em que os preços dos ativos financeiros são imprevisíveis, uma vez que as informações e expectativas são continuamente incorporadas na avaliação dos investidores.

Porém, nas últimas décadas, contrapondo à concepção de mercado eficiente, surgiu uma nova linha de estudos dentro da teoria de finanças, denominado Finanças Comportamentais, que confronta por meio de estudos empíricos o paradigma tradicional baseado no modelo de agente perfeitamente racional proposto por Fama (1970). As Finanças Comportamentais estudam a falibilidade humana nos mercados de capitais (YOSHINAGA et al., 2008). Em consonância com esta nova teoria, Elder (2004), em seu famoso livro *Como se transformar em um operador e investidor de sucesso*, afirma que os investidores não agem sempre racionalmente e com isso a teoria do mercado eficiente só prevaleceria em determinadas condições.

Apesar das dificuldades existentes na previsão de séries temporais financeiras, têm-se empreendido muitos esforços para uma melhor compreensão do mercado acionário. Juntamente com os teóricos das Finanças Comportamentais (SHLEIFER, 2000; BAZERMAN; MOORE, 2008) várias técnicas não convencionais, que utilizam os conceitos de IC aplicados ao mercado de capitais têm questionado a HME e demonstrado que é possível obter lucro no mercado financeiro identificando o melhor momento de comprar e vender ações (ATSALAKIS; VALAVANIS, 2009a; CAVALCANTE et al., 2016; RATHER; SASTRY; AGARWAL, 2017).

1.3 Justificativa

Do ponto de vista estrito da ferramenta desenvolvida, este trabalho de Doutorado se justifica pelos seguintes motivos:

1. Abordagem inédita: os sistemas desenvolvidos utilizam uma abordagem inovadora, que obteve resultados promissores e que, até onde o autor tem conhecimento, é única na literatura.
2. Contesta a HME: o trabalho apresentado contesta a corrente majoritária da Teoria das Finanças, demonstrando que é possível obter lucro no mercado acionário por meio da compra e venda de ações, ignorando assim a HME.
3. Módulo de negociação: os artigos revisados concentram-se principalmente na previsão das séries temporais. Neste trabalho, foi detalhada a forma de efetuar a compra e a venda de ações, não tendo sido encontrado na literatura sistema de negociação semelhante.

4. Estudo sobre a Bolsa de Valores Brasileira: outra contribuição importante são os testes destas implementações com ações da BOVESPA. Nos artigos de revisão Atsalakis e Valavanis (2009a), Cavalcante et al. (2016) e Rather, Sastry e Agarwal (2017) praticamente não são citadas publicações com aplicações na BOVESPA.

1.4 Objetivos

O objetivo desta tese é desenvolver ferramentas direcionadas ao mercado de capitais baseadas em IC e que possam efetuar, automaticamente, a compra e a venda de ações, com o objetivo principal de obter retorno financeiro e evitar perdas financeiras obedecendo aos critérios estabelecidos pelo usuário.

Outro objetivo deste trabalho foi experimentar técnicas que ainda não haviam sido testadas nesse tipo de problema. Foram desenvolvidos dois sistemas de investimento. O Sistema Automatizado de Investimento via Seleção Multiobjetivo de Regras (SisSM) constrói robôs investidores por meio da combinação de regras lógicas provenientes de indicadores da Análise Técnica (AT). Já o Sistema Automatizado de Investimento via Programação Genética Multiobjetivo (SisPG) aprimora os SisSM com a inclusão da Programação Genética (PG) Multiobjetivo para combinação mais eficiente das regras.

1.5 Organização do Trabalho

Este texto é dividido em seis capítulos. No Capítulo 2, é feita uma revisão da literatura, apresentando as técnicas utilizadas para previsão no mercado financeiro e com mais ênfase nos trabalhos que estão no mesmo escopo da proposta desta tese, ou seja, que abordem o tema de PG e AT aplicado ao mercado de capitais. Também são discutidas de forma pontual as diferenças do trabalho em questão com os demais. No Capítulo 3, são discutidas as ferramentas necessárias para construir os dois sistemas automatizados de investimento propostos. São expostos os conceitos de AT, detecção e remoção de *outliers*, seleção de características, PG, otimização mono e multiobjetivo, algoritmos evolucionários multiobjetivo e, por último, é apresentado um modelo de comitê de decisão. No Capítulo 4, são apresentados o SisSM e o SisPG de forma detalhada. No Capítulo 5, são apresentadas a estratégia de teste e os resultados obtidos com os dois sistemas propostos. Foram considerado seis ações da BOVESPA em dois períodos distintos. Por fim, no Capítulo 6, são apresentadas as considerações finais e a proposta de continuidade do trabalho.

2 Revisão Bibliográfica

Este trabalho propõe a divisão das técnicas de previsão no mercado financeiro em três categorias principais:

- métodos do mercado;
- métodos convencionais;
- Sistemas Inteligentes para Previsão de Séries Financeiras (SIPSF).

Os métodos do mercado podem ser divididos em três categorias: a Teoria de Dow, a Análise Técnica (AT) e a Análise Fundamentalista (AF) (MURPHY, 1999; ACHELIS, 2001; ELDER, 1993).

Os métodos convencionais são modelos econométricos que trabalham com a previsão de séries temporais, tais como: o *Autoregressive Moving Average Models* (ARMA), o *Autoregressive Integrated Moving Average* (ARIMA), o *Autoregressive Fractionally Integrated Moving Average* (ARFIMA), o *Seasonal Autoregressive Moving Average* (SARMA), o *Seasonal Autoregressive Integrated Moving Average* (SARIMA), o *Autoregressive Conditional Heteroskedasticity* (ARCH) e o *Generalized AutoRegressive Conditional Heteroskedasticity* (GARCH).

Já os SIPSF, também denominados como métodos não convencionais por Atsalakis (2009b), utilizam técnicas de aprendizagem de máquina para a previsão de séries temporais financeiras.

Este capítulo apresenta a revisão da literatura relacionada aos métodos citados, enfatizando os métodos que utilizam a Programação Genética (PG) aplicada ao mercado financeiro por ser o objeto principal de estudo deste trabalho. Também são apresentadas as contribuições e diferenças dos métodos encontrados na literatura, bem como o trabalho proposto nesta tese.

2.1 Métodos do Mercado

2.1.1 Teoria de Dow

Segundo Murphy (1999), Charle Henry Dow pode ser considerado o pai da Análise Técnica, pois ele definiu os princípios que norteiam o comportamento dos preços dos ativos no mercado financeiro. Charles Dow junto com Edward D. Jones publicaram um informativo financeiro que mais tarde seria o *The Wall Street Journal*. Por meio do jornal, Charles Dow apresentava suas observações sobre o comportamento do mercado. O conjunto desses textos seria posteriormente reunidos, gerando o que pode ser considerado o início da análise técnica. Seus

estudos ficaram conhecidos como a Teoria de Dow e a seguir são apresentados seus princípios básicos (MURPHY, 1999):

1. As médias e os preços descontam tudo: todos os possíveis fatores que determinam o valor dos ativos são descontados por todas as informações do mercado, sejam notícias, resultados contábeis e financeiros, dados de relatórios de companhias, etc..
2. Os mercados realizam três movimentos: as tendências primárias, secundárias e terciárias. As tendências primárias são longos movimentos de alta ou de baixa que duram um ano ou mais e resultam em grandes valorizações ou desvalorizações dos preços. As tendências secundárias são declínios intermediários que ocorrem durante mercados em tendência primária de alta, ou são valorizações que interrompem por algum tempo a queda em um mercado em tendência primária de baixa. As tendências terciárias fazem parte das secundárias. São movimentos menores de, em média, até 3 semanas. Elas se comportam em relação às tendências secundárias da mesma maneira que as secundárias em relação às primárias.
3. A tendência primária é composta por três fases no mercado de alta e três fases no mercado de baixa:
 - Mercado de Alta/Acumulação: as pessoas que detêm informações privadas começam a comprar as ações.
 - Mercado de Alta/Alta: após o mercado começar a subir, começa a se ter um consenso altista, os analistas técnicos conseguem detectar este movimento de alta e iniciam as compras do ativo.
 - Mercado de Alta/Euforia: na mídia só se divulga a alta rentabilidade proporcionada pelo mercado, sendo as projeções dos relatórios financeiros exageradas, e o leigo começa a montar posição de compra. É o início do processo de “bolha”.
 - Mercado de Baixa/Distribuição: os investidores bem informados saem de suas posições vendendo-as aos participantes desinformados. As notícias nos jornais são bastante positivas.
 - Mercado de Baixa/Baixa: depois de algum tempo os preços começam a cair e os investidores mais experientes, junto com os analistas técnicos, começam a se desfazer de suas posições.
 - Mercado de Baixa/Pânico: como a baixa prossegue, não há mais como esconder as evidências, então vem o pânico, o “estouro da bolha”.
4. As tendências devem ser acompanhadas pelo volume: em uma tendência principal de alta é esperado que o volume aumente com a valorização dos ativos e diminua nas reações de desvalorização e em uma tendência principal de baixa é esperado que o volume aumente com a desvalorização dos ativos e diminua nas reações de valorização.

5. A tendência será mantida até que os sinais de reversão sejam confirmados: o mercado não vai cair apenas porque atingiu um nível “alto demais” ou subir porque “já caiu demais”. É necessário que seja identificado um padrão de reversão.

2.1.2 Análise Técnica

“A Análise Técnica é a ciência que, por meio do estudo de registros gráficos dos preços e volumes do passado dos ativos financeiros, busca proporcionar, com base na análise de padrões que se repetem, condições para que possamos projetar o futuro caminho dos preços, dentro de uma lógica de maiores probabilidades” (NORONHA, 1995).

Contrariando a teoria da Hipótese do Mercado Eficiente (HME), que acredita na ineficácia de previsão dos preços de ativos financeiros, os investidores que utilizam a AT entendem que é possível identificar padrões de tendências a partir do comportamento passado de uma série histórica financeira e, desta forma, definir investimentos futuros.

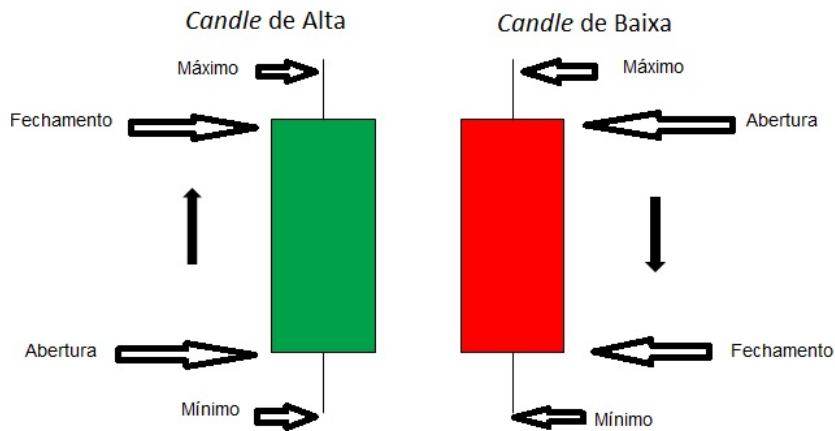
Na AT, a representação gráfica de uma série financeira de um determinado ativo é fundamental. Esta pode indicar uma tendência de alta, uma tendência de baixa ou uma lateralização no preço do ativo ao longo do tempo. Para a construção de um gráfico de uma série financeira que contém a variação diária do preço de um ativo, são utilizadas cinco informações básicas: o preço de abertura do ativo no dia; o preço de fechamento do ativo no dia; o preço mínimo que o ativo atingiu no dia; o preço máximo que o ativo atingiu no dia; e o volume negociado no dia. Essas informações podem ser representadas graficamente por meio de um *Candle*¹, sendo que o conjunto de *Candles* em um determinado intervalo de tempo forma a representação gráfica de uma série financeira. Nas figuras 1 e 2, respectivamente, são apresentados um *Candle* e a variação dos preços de um ativo ao longo de um determinado intervalo de tempo.

Em conformidade com os princípios apresentados na Teoria de Dow, para a AT, a série financeira é um conjunto de informações autocontido, uma vez que os preços descontam tudo (notícias, resultados contábeis, decisões governamentais, etc). Para Murphy (1999), a AT contém três concepções básicas:

- O preço dos ativos representa todos os sentimentos do mercado: o preço diário da ação nada mais é do que uma representação visual das expectativas e comportamento dos investidores.
- O valor dos ativos desloca-se por meio de tendências: o investidor deverá operar a partir do reconhecimento da tendência que o gráfico de uma ação está mostrando.
- O futuro se repete: o mercado tem memória dos acontecimentos passados. Assim, acredita-se que os comportamentos passados tendem a se repetir no futuro.

¹ O *Candle*, também chamado de vela, é uma representação gráfica do que ocorreu com o preço de um ativo no decorrer de um determinado período. Essa forma de representação foi introduzida por japoneses no século XVIII (MURPHY, 1999).

Figura 1 – Componente de um gráfico de *Candles*



Fonte: Elaborada pelo autor

Figura 2 – Exemplo de gráfico de *Candles* aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 30/09/2016 a 01/12/2016. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

A AT pode ser classificada em dois grupos principais: a grafista e a computadorizada (ELDER, 1993). A grafista, também conhecida como clássica, faz uso de reconhecimento de padrões gráficos para determinar a tomada de decisão. Já a computacional faz uso de indicadores numéricos.

2.1.3 Análise Fundamentalista

Esta técnica apresenta outra forma de analisar os investimentos em ativos financeiros. A Análise Fundamentalista (AF) é um método que consiste em estudar os balanços patrimoniais, as margens de lucro, os indicadores da saúde econômica da empresa, avaliando também os aspectos da economia e do setor escolhido para observar possibilidades de aumento no valor da empresa e no preço de sua ação. Este tipo de análise é indicada para investidores de médio e longo prazo, não sendo adequada para investidores de curto e curtíssimo prazo.

Segundo Sá (1999), a AF surgiu nos Estados Unidos, entre os anos de 1920 e 1930, durante o período da crise da Bolsa de Nova Iorque em 1929. Esta crise apontou a necessidade de maior transparência das informações relativas à condição financeira das empresas.

Os seguidores da AF procuram determinar o valor justo das ações de uma determinada empresa, com base na previsão dos lucros e dividendos da empresa em períodos futuros. O objetivo é determinar o valor do ativo através de informações financeiras da empresa.

Para Murphy (1999), a AT e a AF buscam resolver o mesmo problema, que é prever como o preço das ações irá se comportar. A principal diferença entre o Analista Técnico e o Analista Fundamentalista está na forma de observar o mercado. O Analista Fundamentalista tenta responder as seguintes perguntas: “Por que comprar?” e “Por que vender um determinado ativo?”. Por outro lado, o Analista Técnico tem o objetivo de responder outras perguntas: “Quando comprar?” e “Quando vender um determinado ativo?”.

2.2 Métodos Convencionais

Os métodos convencionais são aqueles baseados na econometria. A econometria é um método de análise econômica que agrega a Estatística, a Matemática e a Teoria Econômica.

Entre os objetivos da econometria está a previsão de valores de variáveis econômicas como as séries temporais financeiras, ou seja, esta ferramenta também pode ser usada para prever o valor de ativos financeiros na Bolsa de Valores.

Um modelo econométrico é uma representação simplificada da realidade, que objetiva obter a compreensão parcial ou total de um fenômeno. Em Atsalakis (2009b), é apresentado um amplo estudo sobre estas técnicas. Os modelos econométricos ARMA, ARIMA, ARCH e GARCH fazem parte desse grupo. Segundo Atsalakis (2009b), os modelos ARCH e GARCH apresentaram melhores resultados sobre os demais. As conclusões de Atsalakis (2009b) são confirmadas por Morettin (2011), que destaca o sucesso desses métodos pela capacidade de captar a volatilidade das séries e definir, com maior precisão, a relação entre risco e retorno.

2.3 Sistemas computacionais que fazem uso de inteligência computacional

Nesta seção, a revisão bibliográfica é apresentada em duas partes: a primeira enfatiza os sistemas computacionais que fazem uso de inteligência computacional aplicada ao mercado de maneira geral; já a segunda parte apresenta os sistemas computacionais que fazem uso da PG, que é uma das principais ferramentas utilizadas neste trabalho.

2.3.1 Sistemas Inteligentes para Previsão de Séries Financeiras (SIPSF)

Os SIPSF são sistemas computacionais que fazem uso das mais diversas técnicas da Inteligência Computacional (IC) para prever o preço de um ativo financeiro ou prever um intervalo de compra ou de venda. Esses sistemas utilizam técnicas como redes neurais (CHEN, 1994), *Support Vector Machines* (SVM) (BURGES, 1998), *Particle Swarm Optimization* (PSO) (MAJHI et al., 2008), Lógica *Fuzzy* (DONG; ZHOU, 2002), Algoritmos Genéticos (MARQUES; GOMES, 2009), PG (ALLEN; KARJALAINEN, 1999), entre outros.

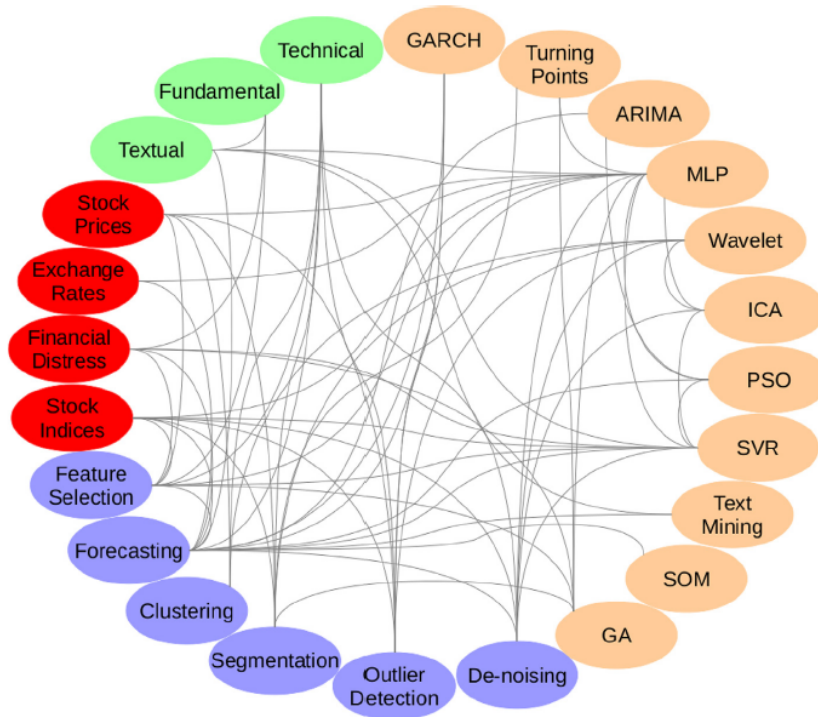
A utilização da IC aplicada ao mercado de capitais é relativamente recente, e teve início na década de 90, com o trabalho pioneiro de Kimoto et al. (1990), que utilizou uma rede neural para fazer previsões na Bolsa de Valores de Tóquio. Desde então, houve um crescimento considerável nesta área com diversas publicações. Apesar da evolução da área, ela ainda carece de métricas universais para classificar e comparar os trabalhos desenvolvidos. Isto acontece principalmente pelo fato dos trabalhos serem avaliados em mercados financeiros distintos, em janelas e frequência de negociações diferentes, não existindo ainda uma base de dados e indicadores globalmente aceitos pela comunidade.

Os trabalhos de *survey* de Atsalakis e Valavanis (2009a), Cavalcante et al. (2016) e Rather, Sastry e Agarwal (2017) auxiliam de alguma forma na classificação das pesquisas realizadas na área. Atsalakis e Valavanis (2009a) fizeram uma ampla investigação de publicações e utilizaram quatro critérios para classificar os trabalhos, tais como: o mercado financeiro que o sistema foi testado; a técnica computacional utilizada; as variáveis de entrada do sistema; o método de validação utilizado. Neste estudo, Atsalakis e Valavanis (2009a) constataram que os métodos computacionais apresentaram melhores resultados que os métodos convencionais apresentados na seção 2.2, pois estes conseguem captar melhor a não estacionaridade das séries financeiras.

Cavalcante et al. (2016) complementaram os trabalhos de Atsalakis e Valavanis (2009a) de duas formas: inclusão de pesquisas mais recentes e consideração de um novo critério de classificação. Foram catalogadas publicações entre 2009 e 2015 utilizando os seguintes critérios: (i) pré-processamento, (ii) previsão e (iii) mineração de texto. Neste trabalho, é apresentado também um diagrama que relaciona os métodos de acordo com as seguintes regras: técnicas de previsão, parâmetros de entrada, áreas de aplicação e objetivos de cada método, como pode ser

observado na figura 3.

Figura 3 – Combinação de técnicas de sistemas computacionais financeiros



Fonte: (CAVALCANTE et al., 2016)

Rather, Sastry e Agarwal (2017) fazem um levantamento histórico dos sistemas de previsão do mercado financeiro dividindo estes em dois grupos: métodos baseados em uma única técnica e métodos híbridos. Neste trabalho, os autores apontam a superioridade dos métodos híbridos em relação aos métodos que utilizam somente uma técnica. Nos trabalhos de revisão de Atsalakis e Valavanis (2009a), Cavalcante et al. (2016) e Rather, Sastry e Agarwal (2017) três observações são destaques:

- a grande maioria dos sistemas de previsão utilizam redes neurais e suas variações, sendo restrito o número de artigos que fazem experimentos com as demais técnicas da IC;
- pouca ênfase nos módulos que tratam a negociação, uma vez que o foco é na previsão das séries financeiras em si, e não como estes dados são utilizados para compras e vendas;
- Atsalakis e Valavanis (2009a) sugerem que os sistemas que de fato estão obtendo retorno financeiro no mercado de ações não estão publicados em periódicos científicos, mas estão guardados como segredo de negócio das instituições financeiras.

2.3.2 A Programação Genética e a Previsão de Séries Financeiras

A PG foi proposta por Koza (1992) e utiliza os conceitos da Computação Evolutiva (CE). A sua estrutura é muito semelhante à de um Algoritmo Genético (AG): inicia-se de uma

população, geralmente aleatória, que é evoluída continuamente, num processo de gerações, por meio de cruzamentos, mutações, e seleção. Contudo, de forma diferente de um AG em que os indivíduos são representados por cadeias numéricas, cada solução candidata em uma PG é representada por uma árvore. Esta representação torna mais fácil a otimização de regras de decisão. Como uma desvantagem, essa estrutura requer um controle adicional sobre cruzamento e mutação para evitar soluções inválidas.

A PG vem sendo aplicada em diversas áreas do conhecimento, como Mineração de Dados, Regressão de Funções, Geração de Regras de Decisão, Reconhecimento de Padrões e Previsão de Séries (ESPEJO; VENTURA; HERRERA, 2010; BARROS et al., 2012) (CORTEZ, 2002; ALFARO-CID; SHARMAN; ESPARCIA-ALCÁZAR, 2014; KATTAN; FATIMA; ARIF, 2015). Atualmente, ela representa uma importante área de estudo dentro de aprendizado de máquina, devido à sua simplicidade e robustez (POLI et al., 2008).

Em diversas áreas do conhecimento ocorrem fenômenos que dependem da observação de dados em determinado intervalo regular de tempo. Esses dados observados no tempo chamam-se séries temporais. O maior desafio do estudo de séries temporais é construir modelos com o menor número de parâmetros possível, de forma a obter previsões precisas e generalizadas. A PG tem sido também utilizada como ferramenta para essa finalidade (CORTEZ, 2002; ALFARO-CID; SHARMAN; ESPARCIA-ALCÁZAR, 2014; KATTAN; FATIMA; ARIF, 2015).

Um caso particular das séries temporais são as séries não lineares e não estacionárias. Dentro desse grupo estão as séries temporais financeiras, que registram a variação do preço de um determinado ativo em relação ao tempo.

A PG também tem sido empregada como técnica de previsão no mercado financeiro. Esta é utilizada de várias maneiras na previsão de séries, seja auxiliando na configuração de pesos de redes neurais, seja na configuração de parâmetros dos Modelos da família ARCH e GARCH ou na construção de árvores de decisão (ALLEN; KARJALAINEN, 1999), (POTVIN; SORIANO; VALLÉE, 2004), (MYSZKOWSKI; RACHWALSKI, 2009), (PIMENTA et al., 2014; VASILAKIS et al., 2013; DABHI; CHAUDHARY, 2015).

Allen e Karjalainen (1999) usaram a PG para geração de regras de negociação. Seus experimentos foram feitos no período de 1928 a 1995 (SP 500 index), utilizando *candles* diários. Para a construção das árvores de decisão foram utilizados Operadores Aritméticos, Máximos, Mínimos e Operadores Condicionais (*if then else*). Os seus experimentos apresentaram resultados positivos, mas não superaram a estratégia *Buy and Hold*.

Vasilakis et al. (2013) usou a PG para a previsão diária da taxa Euro/Dólar. Os seus experimentos foram feitos no período de janeiro de 1999 a outubro de 2009, utilizando *candles* de 15 minutos. Além dos operadores utilizados por Allen e Karjalainen (1999), ele incluiu Operadores de Raiz Quadrada, Raiz Cúbica, Cosseno, Seno, Tangente, Módulo, Logaritmo e Potência. O objetivo era gerar uma expressão matemática para fazer previsões diárias. Fazendo

operações de *daytrade*, seus resultados foram superiores se comparados com três estratégias tradicionais: *Naive*, *MACD* e *Buy and Hold*.

Potvin, Soriano e Vallée (2004) usaram, além dos operadores utilizados por Allen e Karjalainen (1999), alguns indicadores da AT. Os experimentos foram realizados com 14 ações listadas na Bolsa de Toronto, no período de 30 de junho de 1992 a 30 de junho de 2000. Os resultados se apresentaram melhores com ações em queda ou lateralizada. Porém, com o mercado em alta, a estratégia *Buy and Hold* foi a melhor.

Myszkowski e Rachwalski (2009), diferentemente de Allen e Karjalainen (1999) e Potvin, Soriano e Vallée (2004), utilizaram duas árvores de decisão, uma para comprar e outra para vender. Seus experimentos foram realizados no *Forex*² (Euro/Dólar). Foram utilizados com *Candles* de 10 minutos nas operações *daytrade*. As árvores de decisão combinam operadores lógicos (*And*, *Or*) e indicadores da AT. Os resultados se apresentaram instáveis, mas lucrativos para indivíduos com árvores de altura cinco ou seis.

Dabhi e Chaudhary (2015) propuseram um sistema para previsão de série financeira que combina a transformada *wavelet* e a PG com uma notação pós-fixa. A transformada *wavelet* é usada para suavizar a série temporal separando as flutuações da tendência principal. Os testes foram realizados em quatro séries temporais financeiras, dois papéis de empresas (Intel e Microsoft) e dois índices (NASDAQ e S&P CNX). Foram utilizados *Candles* diários, em períodos que variaram de 12 de setembro de 2007 a 11 de novembro de 2010 para (Intel e Microsoft), 1º de março de 2007 a 22 de março de 2011 para (NASDAQ), 1º de março de 1997 a 9 de abril de 2001 para (S&P CNX). O resultado desse trabalho foi comparado com a PG usada em ECJ³ (LUKE et al., 2004). Os testes indicaram que o trabalho proposto foi melhor nos dois índices (NASDAQ e S&P CNX) e não chegaram a uma resposta conclusiva para os papéis de empresas (Intel e Microsoft).

Pimenta et al. (2014) desenvolveu uma proposta semelhante ao trabalho de Myszkowski e Rachwalski (2009). As árvores de decisão geradas combinam operadores lógicos com indicadores da AT. Os experimentos foram realizados no mercado de ações brasileiro - BOVESPA, com *Candles* diários no período de 24 de fevereiro de 2010 a 28 de fevereiro de 2014. O objetivo financeiro estabelecido pelos autores foi obtido em 90% dos casos. A principal crítica a esse trabalho é a não inclusão de limite às perdas financeiras. O trabalho de Pimenta et al. (2014) difere essencialmente de Myszkowski e Rachwalski (2009) em três pontos: o sistema de negociação, o

² FOREX (um acrônimo da expressão em inglês *foreign exchange*, significando Mercado de Câmbio) é um mercado financeiro descentralizado destinado a transações de câmbio, sendo o maior mercado do mundo. Em termos de volume de dinheiro movimentado, movimenta o equivalente a quase 4 trilhões de dólares estadunidenses diariamente. Segundo dados de 2007, movimentava cerca de 3,43 vezes mais do que a soma de todos os mercados de títulos no mundo, e 9,63 vezes o volume negociado no mercado de ações mundial (FOREX, 2014).

³ O ECJ é um sistema de pesquisa de computação evolutiva *freeware* escrito em Java. É uma estrutura que suporta uma variedade de técnicas de computação evolucionária, tais como algoritmos genéticos, programação genética, estratégias de evolução, co-evolução, otimização enxame de partículas, e evolução diferencial.

sistema de verificação e a utilização de meta financeira.

Os sistemas desenvolvidos nessa tese estão na categoria de SIPSF que utilizam a PG aplicada ao mercado de capitais. Além disso, um destes sistemas apresentam semelhanças aos métodos propostos por Potvin, Soriano e Vallée (2004) e Myszkowski e Rachwalski (2009), pois utiliza a PG combinada à AT.

Considerando essas características, a seguir são apresentadas, de forma geral, as principais contribuições dos SIPSF construídos nesse trabalho em relação aos demais:

- combina um método estatístico para a remoção de *Outliers*⁴ da série temporal financeira na fase de treinamento com os seguintes métodos: a Seleção de Características, a PG e o Comitê de Decisão; visando identificar os momentos adequados para a execução das ordens de compra e de venda. Até onde o autor tem conhecimento, tal combinação é única na literatura.
- inclui na etapa de treinamento a PG Multiobjetivo combinada com AT. O arranjo PG e AT já havia sido anteriormente utilizado na literatura, porém, em uma implementação mono-objetivo.
- propõe um sistema de negociação para compra e venda de ações.

No Capítulo 4, são apresentados, de forma detalhada, os sistemas propostos e como essas técnicas são combinadas.

⁴ *Outliers* são valores que parecem ser ou muito grandes ou muito pequenos, em comparação com o restante das observações (GUMBEL, 1960).

3 Ferramentas

O objetivo deste capítulo é desenvolver os conceitos principais das ferramentas utilizadas nos sistemas propostos, sendo: indicadores numéricos da Análise Técnica (AT), Detecção e Remoção de *Outliers*, Seleção de Características, Programação Genética (PG), Algoritmos Evolucionários Multiobjetivos e Comitê de Decisão.

3.1 Os Indicadores Numéricos da Análise Técnica

A AT pode ser definida como

“... o estudo dos preços, volumes e contratos em aberto do mercado, visando prever as tendências futuras de preços”(MURPHY, 1999; ACHELIS, 2001).

Estes métodos são focados na movimentação dos preços em si, e não nas causas que estão por trás dela. Não são consideradas informações de projeção de lucro, de taxa de retorno, de relação preço/lucro e de noticiário em geral.

A AT pode ser dividida em dois grupos: a clássica, que faz uso principalmente de reconhecimento de padrões gráficos; e, a computadorizada (ELDER, 1993), que faz uso de indicadores numéricos.

Nessa secção são apresentados os indicadores numéricos da AT, os quais são utilizados nos sistemas propostos nessa tese. Ao todo, podem ser contabilizados mais de cem indicadores numéricos que podem ser divididos em quatro categorias principais (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001):

- **Seguidores de tendência:** os indicadores dessa categoria identificam a movimentação principal dos preços de um ativo em um determinado período. Como exemplo, podem ser citados a Média Móvel Simples (MMS), a Média Móvel Exponencial (MME), os canais de *Donchian* e o ativador *Hilo*.
- **Osciladores:** os indicadores dessa categoria monitoram as variações do preço de um ativo em uma determinada faixa com o objetivo de sinalizar possíveis pontos de reversão dos movimentos de alta ou de baixa. Como exemplo, podem ser citados o Oscilador de Chaikin (OC), a Volatilidade do Oscilador de Chaikin (VC) e o Williams %R (WR).
- **Sistemas de Banda:** os sistemas de banda são constituídos por três curvas relacionadas aos preços. Elas são traçadas a partir de uma determinada distância de uma média móvel. A banda intermediária é uma medida de tendência mediana, geralmente uma média móvel

simples, que serve como base para as bandas inferiores e superiores. Os intervalos entre a banda intermediária e as bandas superior e inferior são definidos com base na volatilidade (geralmente o desvio padrão das mesmas datas que utilizadas para a média). Assim, quanto maior a volatilidade de um ativo, maior seu desvio padrão. Quando não há tendência definida, a regra é vender quando o preço ficar acima da banda superior e comprar quando o preço ficar abaixo da banda inferior. Como exemplo, podem ser citados os canais de *Keltner*, as Bandas de Bollinger (BB) e os osciladores *Bollinger*.

- **Identificadores de divergência:** estes indicadores se baseiam no princípio de que toda a tendência passa por correções. As divergências ocorrem quando se compara o comportamento do indicador em relação ao movimento do preço de um ativo. Como exemplo, podem ser citados a Acumulação e Distribuição (AD), o Saldo de Volume (OBV) e o Índice de Força Relativa (IFR).

Na tabela 1, são apresentados alguns dos principais indicadores da AT (MURPHY, 1999; ELDER, 1993). Na sequência, são detalhados cada um dos indicadores técnicos utilizados nos sistemas propostos.

3.1.1 Média Móvel Simples (MMS)

A MMS fornece o valor médio da cotação de preço de uma ação em um determinado período, sendo que cada dado utilizado no cálculo da média terá o mesmo peso (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). A MMS segue uma tendência, uma vez que ela não antecipa mudanças no comportamento dos preços. O número de períodos considerados para construir uma MMS fica a cargo do investidor. A direção de sua inclinação indica a tendência do mercado, de modo que, quando inclinada para cima, sugere um mercado em alta, e quando inclinada para baixo, sugere um mercado em baixa, como pode ser observado na figura 4. A sua equação é apresentada em 3.1.

$$MMS_w(i) = \frac{\sum_{u=(i+1-w)}^i P_{fu}}{w} \quad (3.1)$$

Onde:

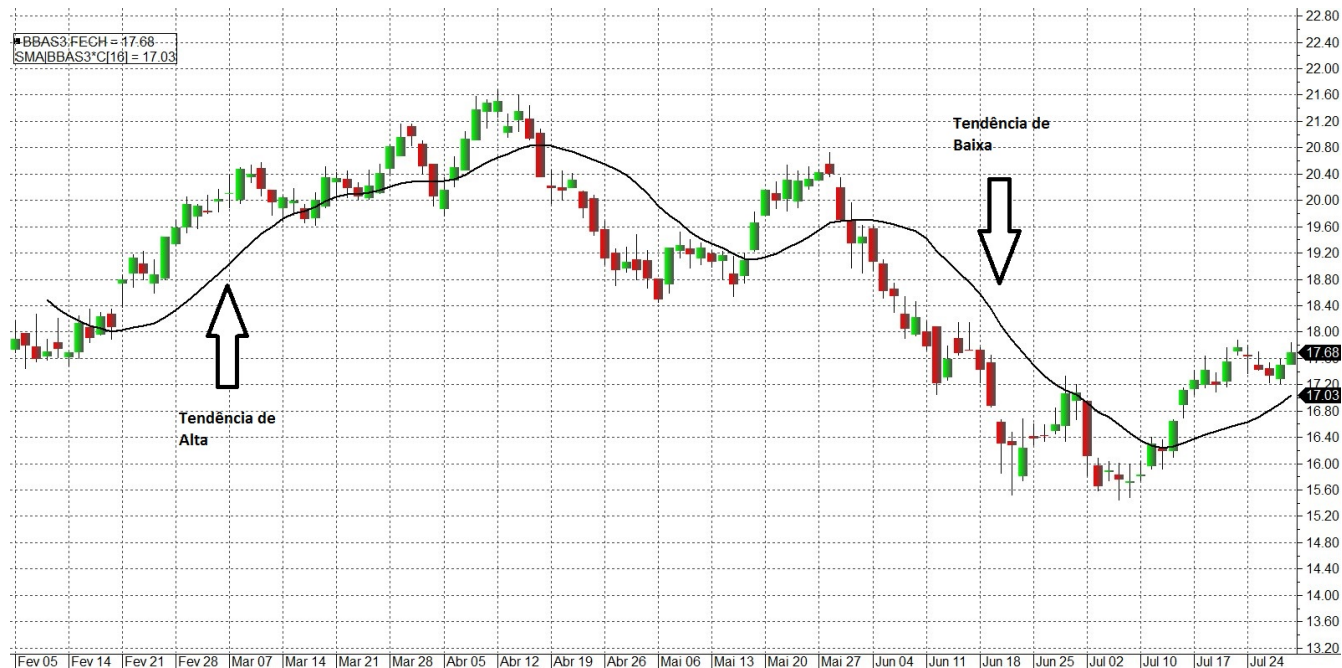
- i : período corrente (data atual);
- w : número de períodos;
- P_f : preço de uma ação no fechamento de um período.

Tabela 1 – Principais indicadores técnicos

Id	Nome
1	Média Móvel Adaptativa
2	Média Móvel Exponencial
3	Média Móvel Simples
4	Média Móvel Ponderada
5	Média Móvel Tripla
6	Bandas Bollinger
7	Canais Donchian
8	Médias Móveis Envelopes
9	Diferença Compra-Venda
10	Bandas de distorção (SkewBands)
11	CCI
12	CCI/MA cruzamento
13	SAR Parabólico
14	MACD
15	Histograma MACD
16	Regressão Linear
17	Desvio Padrão
18	Volume
19	Volume+
20	Volume AMA
21	Volume EMA
22	Volume SMA
23	Acumulação de Volume
24	Eficiência do Volume
25	Momentum
26	Índice de Força Relativa (IFR)
27	Estocástico
28	Oscilador de Momento de Chande
29	Curva de Coppock
30	Índice de Massa
31	Fluxo de Dinheiro
32	Saldo de Volume (On-balance Volume)
33	Volatilidade
34	Indicador de Volatilidade de Chaikin
35	Curtose
36	Curtose rápida/lenta
37	Média da amplitude de variação
38	Distorção
39	Oscilador de Chaikin
40	Oscilador de Preço Detrended

Fonte: Elaborada pelo autor

Figura 4 – Exemplo de MMS aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

3.1.2 Média Móvel Exponencial (MME)

A MME é uma extensão da MMS que pondera as observações passadas, dando maior peso aos últimos valores do cálculo (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). A MME geralmente reage com mais rapidez às mudanças do mercado. A direção de sua inclinação também indica a tendência do mercado, conforme pode ser observado na figura 5. A sua equação é apresentada em 3.2.

$$MME_w(i) = P_{fi} \times k + MME_w(i-1) \times (1-k) \quad (3.2)$$

Onde:

- i : período corrente (data atual);
- w : número de períodos;
- z : $(i + 1 - w)$;
- $MME_w(z)$: $MMS_w(z)$;
- $MMS_w(z)$: equação 3.1;

- $k: \frac{2}{w+1}$;
- P_f : preço de uma ação no fechamento de um período.

Figura 5 – Exemplo de MME aplicada na série histórica das ações do Banco do Brasil(BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

3.1.3 Acumulação e Distribuição (AD)

O indicador de AD associa as mudanças no preço e no volume. Quanto maior o volume que acompanha uma mudança no preço, mais significativa será essa mudança (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). Divergências entre o indicador AD e o preço da ação sugerem que haverá uma mudança na tendência, como pode ser observado na figura 6. A sua equação é apresentada em 3.3.

$$AD(i) = Vol_i \times \frac{(P_{fi} - P_{min(i)}) - (P_{max(i)} - P_{fi})}{P_{max(i)} - P_{min(i)}} \quad (3.3)$$

Onde:

- i : período corrente (data atual);
- Vol : quantidade de ações negociadas em um período;

- P_f : preço de uma ação no fechamento de um período;
- P_{min} = preço mínimo alcançado por uma ação em um período;
- P_{max} = Preço máximo alcançado por uma ação em um período.

Figura 6 – Exemplo de AD aplicada na série histórica das ações do Banco do Brasil(BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

3.1.4 Oscilador de Chaikin (OC)

Os osciladores ajudam a identificar pontos de inflexão e permitem que se encontrem níveis insustentáveis de altas e de baixas no mercado acionário (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). O OC flutua em uma escala entre valores positivos e negativos e a sua fórmula é feita através da diferença entre uma MME de 3 períodos do indicador de AD e uma MME de 10 períodos do indicador de AD, como apresentado na equação 3.4. O indicador fica no campo positivo quando há forte pressão compradora e fica no campo negativo quando há pressão vendedora sobre algum ativo, como pode ser observado na figura 7.

$$OC(i) = (MME_3(AD(i)) - MME_{10}(AD(i))) \quad (3.4)$$

Onde:

- i : período corrente (data atual);

- *MME*: equação 3.2;
- *AD*: equação 3.3.

Figura 7 – Exemplo de OC aplicado em uma série financeira



Fonte: Gráfico retirado do site (Técnicas de Trading) (TÉCNICASTRADING, 2017)

3.1.5 Volatilidade do Oscilador de Chaikin (VC)

Este indicador avalia a diferença entre os preços máximos e mínimos de uma determinada ação e usa essa diferença como um indicador de volatilidade (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). Existem duas formas de interpretar essa volatilidade: a primeira assume que os topos do mercado são quase sempre acompanhados por um aumento da volatilidade e os fundos de um mercado são geralmente acompanhados pela queda da volatilidade; o segundo método assume que um aumento na volatilidade, sobre um período de tempo relativamente curto, indica que um fundo está próximo e que uma diminuição da volatilidade sobre um período de

tempo mais longo indica uma aproximação do topo, como pode ser observado na figura 8. A sua equação é apresentada a seguir, na equação 3.5.

$$VC(i) = 100 \times \frac{MME_{10}(P(i)_{max} - P(i)_{min}) - MME_{10}(P(i-10)_{max} - P(i-10)_{min})}{MME_{10}(P(i)_{max} - P_{10}(i)_{min})} \quad (3.5)$$

Onde:

- i : período corrente (data atual);
- MME : equação 3.2;
- P_{min} = preço mínimo alcançado por uma ação em um período;
- P_{max} = preço máximo alcançado por uma ação em um período.

Figura 8 – Exemplo de VC aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

3.1.6 Williams %R (WR)

O indicador WR compara o preço de fechamento de um ativo com relação ao valor mais alto e mais baixo dos últimos w períodos. Ele funciona como um oscilador estocástico (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). Os seus valores variam de -100 a 0 e

ele indica se um ativo está sobrecomprado ou sobrevendido. Geralmente assume-se que um valor abaixo de -80 indica que o ativo está sobrevendido, e acima de -20 indica que este está sobrecomprado, como pode ser observado na figura 9. A sua equação é apresentada na equação 3.6.

$$WR_w(i) = \frac{P_f(i) - H_w}{H_w - L_w} \times 100 \quad (3.6)$$

Onde:

- i : período corrente (data atual);
- w : número de períodos;
- P_f : preço de uma ação no fechamento de um período;
- L_w = preço de venda mais baixo dos últimos w períodos;
- H_w = preço de compra mais alto dos últimos w períodos.

Figura 9 – Exemplo de WR aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

3.1.7 Volume Taxa de Variação (VROC)

O indicador VROC consegue captar a variação na quantidade de ações negociadas (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). Um aumento nos preços, juntamente com

um aumento no volume de ações negociadas, pode indicar uma tendência de valorização da ação. De maneira inversa, um aumento no volume de ações negociadas e uma redução nos preços podem indicar uma queda na cotação das ações. O indicador é a diferença entre o volume atual e o volume de w períodos atrás. Esse indicador é ilustrado na figura 10 e definido na equação 3.7.

$$VROC_w(i) = \frac{(Vol(i) - Vol(i - w))}{Vol(i - w)} \times 100 \quad (3.7)$$

Onde:

- i : período corrente (data atual);
- w : número de períodos;
- $Vol(i)$: volume negociado no período corrente.

Figura 10 – Exemplo de VROC aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

3.1.8 Preço Taxa de Variação (PROC)

O indicador PROC ajuda a identificar uma tendência, pois ele mede a diferença de preços em um determinado período, refletindo esse movimento ondulatório como um oscilador. O PROC cresce com o aumento nos preços e decresce com a queda nos preços. Quanto maior

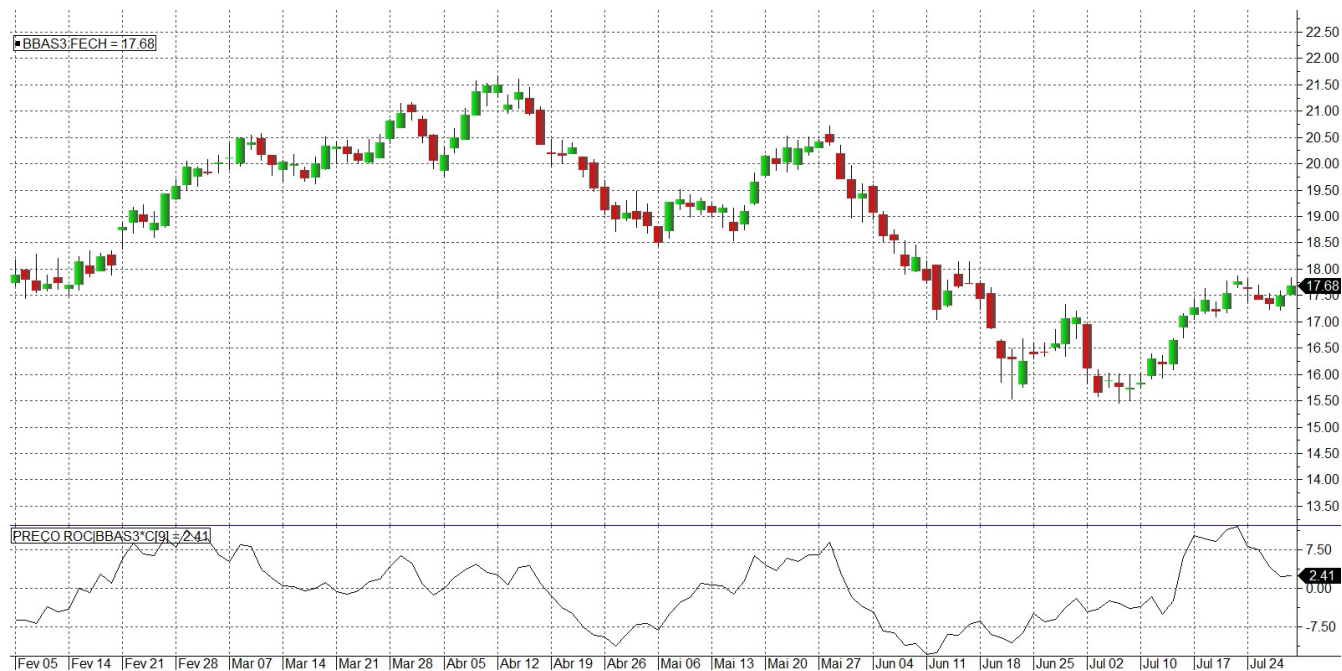
for a variação nos preços, maior será a variação em PROC (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). Quando o oscilador de preços varia positivamente, isso indica uma tendência de alta; quando o oscilador de preços varia negativamente, isso indica uma tendência de baixa. Este indicador é ilustrado na figura 11 e definido na equação 3.8.

$$PROC_w(i) = \frac{(P_f(i) - P_f(i - w))}{P_f(i - w)} \times 100 \quad (3.8)$$

Onde:

- i : período corrente (data atual);
- w : número de períodos;
- P_f : preço de uma ação no fechamento de um período.

Figura 11 – Exemplo de PROC aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

3.1.9 Bandas de Bollinger (BB)

As BB são construídas por três linhas: a linha central (*BBC*), a linha superior (*BBS*) e a linha inferior (*BBI*). A linha central é uma MMS, a linha superior é a linha central mais duas vezes o seu desvio padrão, e a linha inferior é igual à linha central menos duas vezes o seu desvio

padrão (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). O desvio padrão representa o nível de volatilidade do ativo. Em períodos em que não se tem uma tendência definida, o valor de um ativo acima da linha superior indica uma venda. Por outro lado, quando o valor do ativo está abaixo da linha inferior, isto indica uma compra. As linhas das BB (*BBS*, *BBC* e *BBI*) são definidas nas equações 3.9, 3.10 e 3.11, e ilustradas na figura 12.

$$BBS_w(i) = MMS_w(i) + (2 \times \sigma) \quad (3.9)$$

$$BBC_w(i) = MMS_w(i) \quad (3.10)$$

$$BBI_w(i) = MMS_w(i) - (2 \times \sigma) \quad (3.11)$$

Onde:

- i : período corrente (data atual);
- w : número de períodos;
- σ : desvio padrão dos preços de w dias;
- $MMS_w(i)$: definido na equação 3.1.

3.1.10 Saldo de Volume (OBV)

O indicador OBV utiliza o conceito de que o volume precede o preço da ação e é utilizado para verificar o fluxo positivo ou negativo do volume (MURPHY, 1999; ELDER, 1993; ACHELIS, 2001). O OBV é um indicador de momento e relaciona o volume com as mudanças de preços. Divergências entre o OBV e o preço da ação sugerem que haverá uma mudança na tendência. O saldo de volume é calculado por meio da adição do volume do período quando o preço de fechamento do ativo for positivo e da subtração do volume do período quando o preço de fechamento do ativo for negativo. Esse indicador é ilustrado na figura 13 e definido nas equações 3.12, 3.13 e 3.14.

Para calcular o OBV, basta seguir os seguintes passos:

1. se o preço da ação fecha o período em baixa, o valor do volume do período é subtraído do indicador:

$$OBV(i) = OBV(i - 1) - Vol(i) \quad (3.12)$$

Figura 12 – Exemplo de BB aplicadas na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos(XP, 2017).



Fonte: Elaborada pelo autor.

2. se o preço da ação fecha o período em alta, o valor do volume do período é somado ao indicador:

$$OBV(i) = OBV(i - 1) + Vol(i) \quad (3.13)$$

3. se o preço da ação é estável durante o período, o volume do período não é adicionado e nem subtraído do indicador:

$$OBV(i) = OBV(i - 1) \quad (3.14)$$

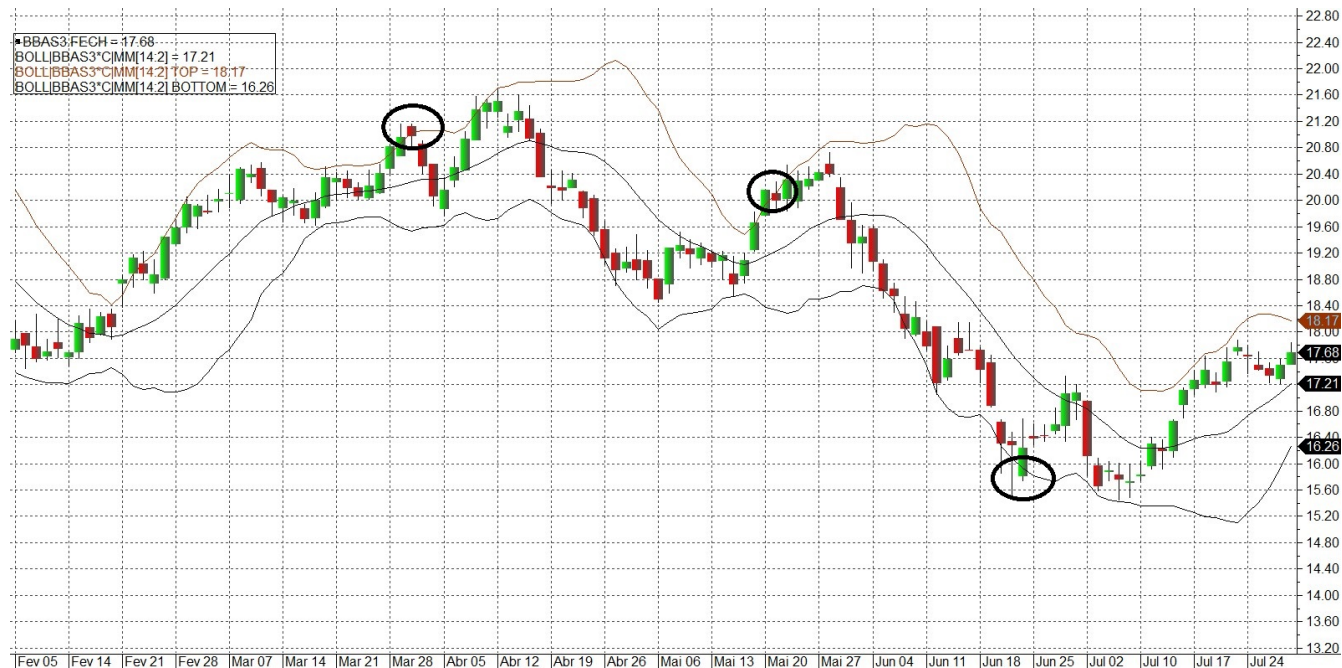
Onde:

- i : período corrente (data atual);
- $Vol(i)$: é o volume corrente.

3.1.11 Índice de Força Relativa (IFR)

O IFR identifica a evolução da relação de forças entre compradores e vendedores ao longo do tempo. Sua utilização possibilita observar o enfraquecimento de uma tendência, rompimentos, suporte e resistência antes de se tornarem aparentes no gráfico de barras (MURPHY, 1999; ELDER, 1993). Seus valores podem variar entre zero e cem. Valores acima de 80 indicam

Figura 13 – Exemplo de OBV aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

que o ativo está sobrecomprado. Já valores abaixo de 20 indicam que o ativo está sobrevendido. Esse indicador é ilustrado na figura 14 e definido na equação 3.15.

$$IFR_w(i) = \frac{100}{1 + \frac{D_w(i)}{U_w(i)}} \quad (3.15)$$

Onde:

- i : período corrente (data atual);
- w : número de períodos;
- U_w : média de todas as variações positivas no preço da ação, dentro do período em estudo w ;
- D_w : média de todas as variações negativas no preço da ação, dentro do período em estudo w .

3.1.12 Estocástico (E)

O indicador E indica a relação entre o preço de fechamento e o preço máximo/mínimo do período de tempo atual. Dois osciladores estocásticos são geralmente calculados para estimar

Figura 14 – Exemplo de IFR aplicada na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

variações futuras nos preços: o rápido (%K) e o devagar (%D). Comparações entre eles são um bom indicador referente à velocidade com que os preços estão mudando (MURPHY, 1999; ELDER, 1993). Esse indicador é ilustrado na figura 15 e definido nas equações 3.16 e 3.17.

$$E_K(i) = 100 \times \frac{P_f(i) - VMax_K}{VMax_K - VMin_K} \quad (3.16)$$

Onde:

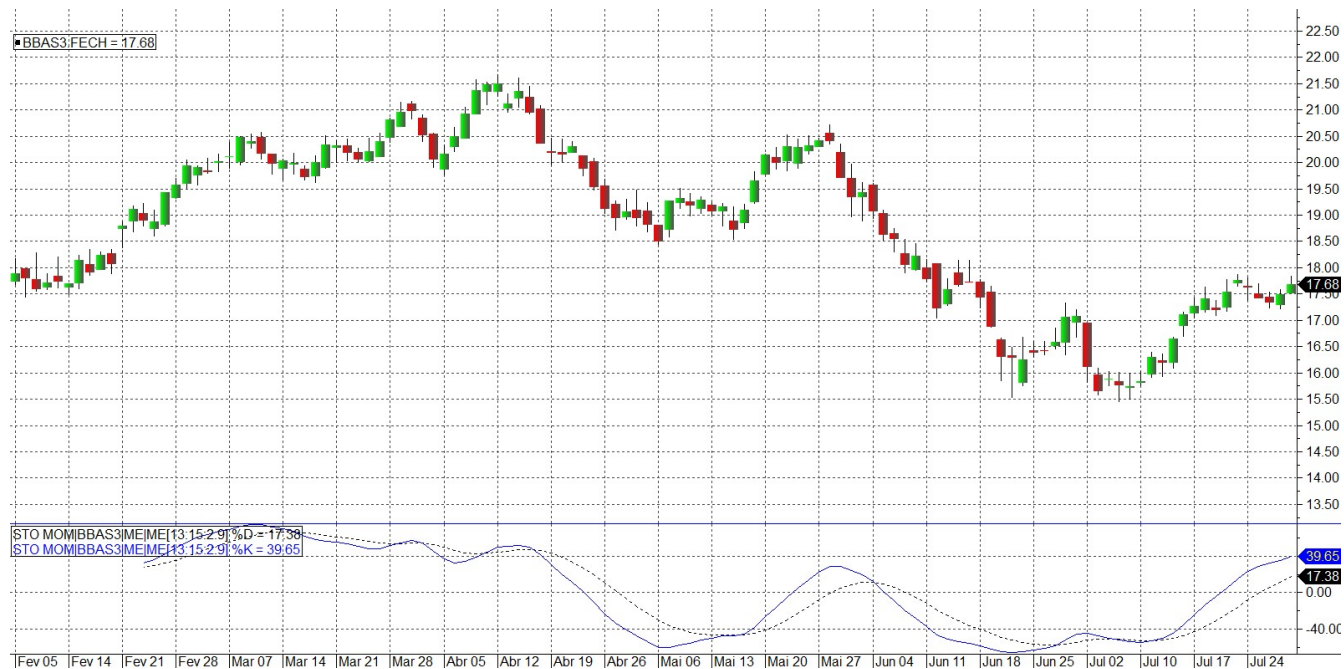
- i : período corrente (data atual);
- P_f : preço de uma ação no fechamento de um período;
- K : número de períodos;
- $VMax_K$: valor mais alto em k períodos;
- $VMin_K$: valor mais baixo em k períodos.

$$E_D(i) = MMS(E_k, D) \quad (3.17)$$

Onde:

- i : período corrente (data atual);
- $MMS(X, Y)$: média móvel simples de X em Y períodos.

Figura 15 – Exemplo de Estocástico (E) aplicado na série histórica das ações do Banco do Brasil (BBAS3), entre o período de 05/02/2013 a 30/07/2013. Gráfico gerado pelo aplicativo XP PRO Versão 1.26.13 e distribuído para clientes da empresa XP-Investimentos (XP, 2017)



Fonte: Elaborada pelo autor

3.2 Detecção e Eliminação de *Outliers*

As séries temporais financeiras são influenciadas diariamente pelos mais diversos fatores, tais como alteração cambial, mudanças econômicas ou políticas, greves e notícias falsas. As ocorrências destas manifestações podem gerar a presença de observações inconsistentes das demais observações da série, denominadas *Outliers*. Os *Outliers* podem ser interpretados como valores atípicos em um conjunto de dados, sejam esses muito altos ou muito baixos (GUMBEL, 1960; GRUBBS, 1969; HAWKINS, 1980; BARNETT; LEWIS, 1994).

Esses valores, por serem atípicos, podem atrapalhar de forma significativa a regressão de modelos que representem adequadamente o comportamento de séries financeiras. Logo, para atenuar esses problemas é necessário detectá-los e eliminá-los. A detecção de *outliers* em séries temporais foi iniciada por Fox (1972). Nos dias de hoje, existem vários métodos para detectar e remover os *outliers* em séries temporais, tais como: (i) o método do Quartil (TUKEY, 1977);

(ii) o método do Quartil ajustado (HUBERT; VANDERVIEREN, 2008) e (iii) método de cercas resistentes (THOMPSON; SIGMAN, 1996).

Dentre os vários métodos estatísticos de detecção de *Outlier* citados, o método do Quartil proposto por Tukey (1977) é o mais conhecido, sendo os demais geralmente derivações deste. Não é objeto de estudo desta tese uma ampla exploração de métodos de detecção de *Outliers*. A seguir, é apresentado o método de Tukey (1977). Para informações sobre os demais métodos é recomendado consultar as referências originais.

3.2.1 Método do Quartil

Na estatística descritiva, as separatrizes são valores que dividem a distribuição em um certo número de partes iguais: a mediana divide em 2 partes iguais, os quartis dividem em 4 partes iguais e os decis em 10 partes iguais. O objetivo das separatrizes é proporcionar uma melhor ideia da dispersão do conjunto, principalmente da simetria ou assimetria da distribuição (LARSON; FARBER; PATARRA, 2004).

Os quartis são as separatrizes que dividem o conjunto em 4 partes iguais. O primeiro quartil (Q_1) é o valor do conjunto que delimita os 25% menores valores: 25% dos valores são menores do que (Q_1) e 75% são maiores do que (Q_1). O segundo quartil (Q_2) separa os 50% menores dos 50% maiores valores. O terceiro quartil (Q_3) é o valor que delimita os 25% maiores valores: 75% dos valores são menores do que (Q_3) e 25% são maiores do que (Q_3) (LARSON; FARBER; PATARRA, 2004).

O Método do Quartil utiliza o conceito de intervalo de tolerância: valores pertencentes a este intervalo são considerados aceitáveis, enquanto que valores fora deste (abaixo ou acima) são tratados como *Outliers*. O intervalo de tolerância gerado pelo Método do Quartil é apresentado na equação 3.18, onde Q_1 e Q_3 são, respectivamente, o primeiro quartil e o terceiro quartil, IQR é o intervalo interquartil, dado por $Q_3 - Q_1$, e C é uma constante que pode ser ajustada de acordo com a distribuição.

$$[Q_1 - C \times IQR, Q_3 + C \times IQR] \quad (3.18)$$

A seguir é demonstrado um exemplo com o objetivo de identificar os possíveis *Outliers* de um conjunto de números ordenados apresentados na tabela 2. Para tanto, foi utilizado o Método do Quartil considerando a constante C igual a 1.5.

Tabela 2 – Conjunto de números inteiros ordenados

2	6	9	12	14	15	17	18	41
1º Quartil			2º Quartil		3º Quartil		4º Quartil	

Fonte: Elaborada pelo autor

Seguem as etapas:

1. Identificação de Q_1 e Q_3 . Considerando que 25% dos valores do conjunto são menores do que Q_1 e 25% dos valores do conjunto são maiores que Q_3 , logo Q_1 e Q_3 são respectivamente o número 9 e o número 17;
2. Cálculo do IRQ . $IRQ = Q_3 - Q_1$, logo, IRQ é igual a 8;
3. Cálculo do intervalo de tolerância utilizando a equação 3.18 e considerando que Q_1 , Q_3 , IRQ e C são respectivamente 9, 17, 8 e 1.5, o intervalo de tolerância é $[-3, 39]$.

Portanto, para o conjunto de números apresentados na tabela 2, utilizando o método do Quartil, o número 41 é considerado um *outlier*.

No capítulo 4, subseção 4.3.3, é detalhado o algoritmo de detecção e remoção de *outlier* utilizado nesse trabalho.

3.3 Seleção de Características

A Seleção de Características objetiva minimizar o problema da “Maldição da Dimensionalidade”(DUDA; HART; STORK, 2012), cuja questão central está no aumento exponencial do espaço de busca para as tarefas de Aprendizagem de Máquina (AM). O fenômeno da “Maldição da Dimensionalidade” aumenta a complexidade nas atividades relacionadas à AM, degradando o desempenho dos algoritmos de exploração e análise de dados. Uma maneira de reduzir estes problemas é por meio da escolha das características mais significativas, utilizando técnicas para tal fim.

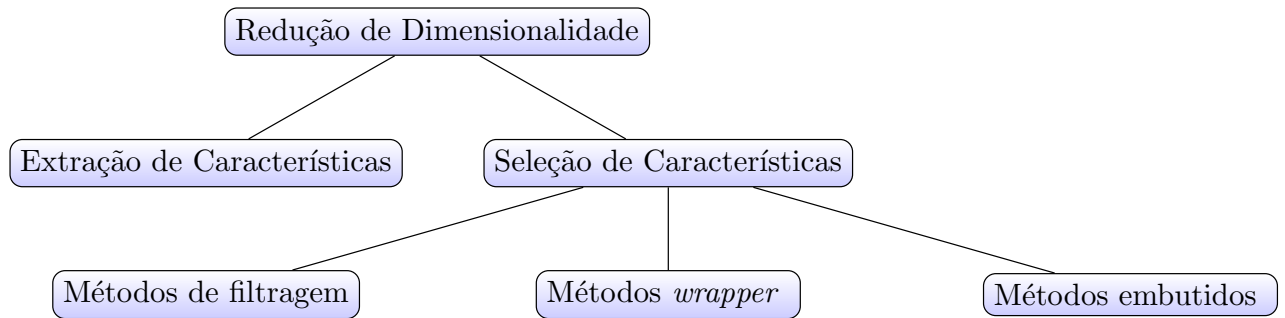
3.3.1 Redução da Dimensionalidade

Existem duas classes de métodos de Redução de Dimensionalidade: a Seleção de Características e a extração de características. A Seleção de Características busca encontrar o subconjunto de características mais relevantes do conjunto de dados originais de acordo com uma regra de avaliação. A extração de características atua de outra forma, projetando as características de seu espaço inicial para um novo espaço de menor dimensionalidade, não mantendo as dimensões originais (DUDA; HART; STORK, 2012). Na figura 16, é apresentada a classificação dos métodos de Redução de Dimensionalidade.

3.3.2 Classes de métodos de seleção de características

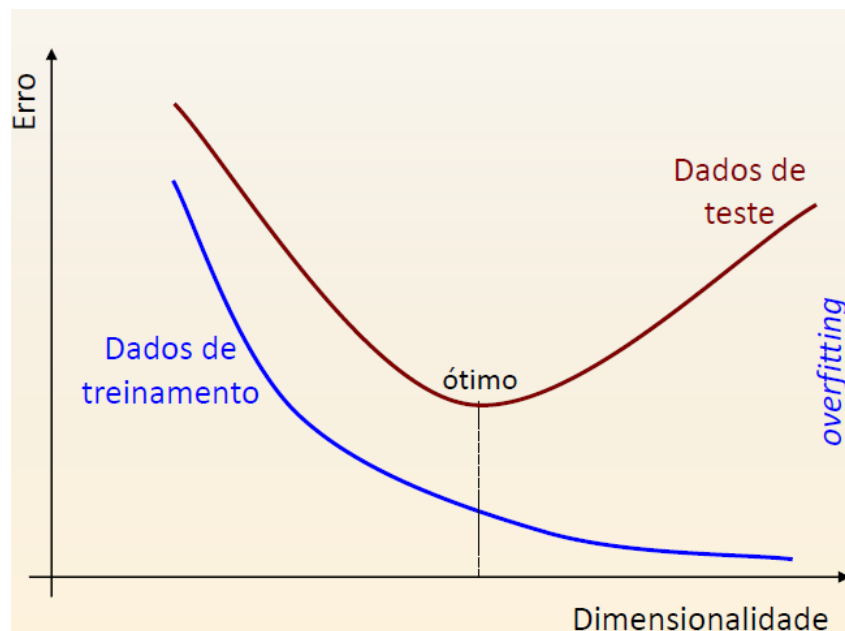
Como apresentado na figura 16, os métodos de seleção de características podem ser classificados em três grupos principais (DUDA; HART; STORK, 2012) de acordo com o critério de avaliação empregado:

Figura 16 – Técnicas de Redução de Dimensionalidade



Fonte: Elaborada pelo autor

Figura 17 – *Overfitting* em função da dimensionalidade



Fonte: Silva (2011)

1. Métodos de Filtragem;
2. Métodos *Wrapper*;
3. Métodos Embutidos.

Os métodos de filtragem estimam as características individuais baseando-se em propriedades intrínsecas dos dados, sem envolver qualquer algoritmo de aplicação. Estes métodos são escaláveis mesmo em conjuntos de dados de dimensionalidade elevada.

Os métodos *wrapper* avaliam um subconjunto de características com base no desempenho de um classificador. Geralmente o classificador utilizado na avaliação do subconjunto é o mesmo utilizado na aplicação.

Os métodos embutidos integram a seleção de características ao algoritmo de aplicação, geralmente um classificador. A relevância de características é avaliada com base na sua utilidade para a otimização da função objetivo do modelo de inferência. Desta forma, a busca pelas características mais relevantes é guiada pelo processo de aprendizagem.

No capítulo 4, subseção 4.3.4, é descrito o algoritmo de seleção de características utilizado neste trabalho.

3.4 Programação Genética

A CE tem como base a Teoria da Evolução das Espécies de Charles Darwin (HOWARD, 1982), e é composta por algoritmos de busca baseados nos princípios da seleção natural. A PG, que é uma das técnicas da CE, foi desenvolvida por Koza (1992) e tem como principal característica o uso de indivíduos que representam códigos de computador. A PG pode ser utilizada para resolver problemas que tenham soluções que possam ser representadas utilizando a sintaxe de uma linguagem de programação e que possam ser avaliadas por uma função de avaliação. Essa função é utilizada para medir a qualidade de um programa computacional para atender uma determinada solicitação. O pseudocódigo 1 mostra os principais passos de um algoritmo de PG.

Algorithm 1 Pseudocódigo do algoritmo PG

- 1: Gera-se a População Inicial P .
 - 2: Avalia-se os indivíduos de P por uma a função de aptidão.
 - 3: Selecionam-se os indivíduos mais promissores.
 - 4: Aplicam-se os operadores genéticos de mutação e cruzamento com certa probabilidade.
 - 5: Adicionam-se em P os novos indivíduos de acordo com o método de reprodução.
 - 6: Se o critério de parada não é atendido, retorna-se ao passo 2.
-

A PG utiliza como critério de parada: número máximo de gerações, tempo limite de execução, aptidão atinge um limite aceitável, dentre outros.

Nas subseções seguintes são descritos de maneira concisa, a forma de representação dos programas, as técnicas de geração da população inicial, a avaliação dos indivíduos, os métodos de seleção e os operadores genéticos.

3.4.1 Representação dos Programas

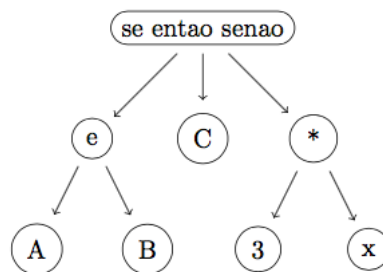
Na PG, os programas de computador podem ser representados por diferentes estruturas: linear (OLTEAN et al., 2009), árvore (KOZA, 1992) ou grafos (TELLER, 1996). Dentre essas estruturas, a representação por árvores é a mais utilizada, pois esta facilita a aplicação dos operadores de mutação e cruzamento. Nesta representação, o cromossomo de cada indivíduo é formado por uma combinação de elementos dos conjuntos Funções (\mathbb{F}) e Terminais (\mathbb{T}), que

variam de acordo com o domínio da aplicação. As funções do conjunto \mathbb{F} são os operadores, ou seja, funções que exigem um determinado número de argumentos de entrada. Os operadores são classificados em operadores matemáticos, lógicos, relacionais, condicionais, dentre outros. Já o conjunto \mathbb{T} contém os operandos, ou seja, variáveis ou constantes. Esses conjuntos devem ser consistentes no sentido de que cada operador de \mathbb{F} aceite como parâmetros: os terminais de \mathbb{T} ou retornos de funções de \mathbb{F} de um determinado tipo (AUGUSTO; BARBOSA, 2013).

A Figura 18 mostra a árvore associada a um indivíduo da população que representa a expressão:

se A e B então C , senão $3 * X$.

Figura 18 – Representação de um indivíduo pela estrutura árvore



Fonte: (AUGUSTO; BARBOSA, 2013)

3.4.2 Geração da População Inicial

A primeira etapa da PG é a Geração da População Inicial P_1 . Essa população deve ser criada de modo que ela seja representativa do espaço dos programas, bem distribuída e não favoreça à geração de um particular conjunto de programas. Os indivíduos de P_1 são gerados de forma aleatória a partir de combinações recursivas de elementos de \mathbb{F} e de \mathbb{T} respeitando o número de argumentos de cada operador em \mathbb{F} . Os indivíduos gerados podem apresentar diferentes níveis de complexidade.

Existem vários métodos para gerar P_1 , sendo os mais comuns: o *Full*, o *Grow* e o *Ramped Half-and-Half* (KOZA, 1992). No método de geração *Full*, as árvores são construídas de modo que a distância entre o nó-raiz e qualquer nó terminal seja igual a uma profundidade específica. Já o método *Grow* permite a geração de árvores que possuam dois nós terminais localizados a diferentes distâncias do nó-raiz limitadas a uma profundidade máxima. O método *Ramped Half-and-Half* define um conjunto de segmentos de modo que cada um tenha uma quantidade fixa de indivíduos e uma profundidade máxima definida. Em seguida, para cada profundidade máxima gera-se metade dos indivíduos utilizando o método *Full* e, a outra metade, utilizando-se o método *Grow*. Segundo Koza (1992), esse método aumenta a diversidade das soluções.

3.4.3 Avaliação dos Indivíduos

Uma função de avaliação é a maneira utilizada pela PG para medir a aptidão de um indivíduo, ou seja, a qualidade de um programa para realizar determinada tarefa (LINDEN, 2008). Segundo Augusto e Barbosa (2013), os programas são executados para um conjunto de argumentos e a aptidão de cada programa é medida comparando-se seu valor de retorno com um valor previamente conhecido. Quanto mais próximo ao valor esperado, maior a aptidão de um programa.

Em problemas de classificação, o objetivo é medir a qualidade da predição para uma amostra de treinamento. Desse modo, pode-se definir para um indivíduo sua função de aptidão com base na taxa de acertos. Outro modo de avaliar um indivíduo é ponderar sua aptidão em função de sua complexidade ou tamanho, com o objetivo de obter soluções mais compactas e com maior capacidade de generalização.

3.4.4 Métodos de Seleção

Um Método de Seleção atua como uma maneira de simular o mecanismo de seleção natural, onde os mais aptos têm maiores chances de sobreviver e gerar descendentes com suas características. O conceito fundamental é aumentar as chances de “reprodução” dos indivíduos com maiores valores de função. Existem várias técnicas de seleção, como (KOZA, 1992):

- Roleta;
- Truncamento;
- Amostragem Estocástica Uniforme;
- Torneio.

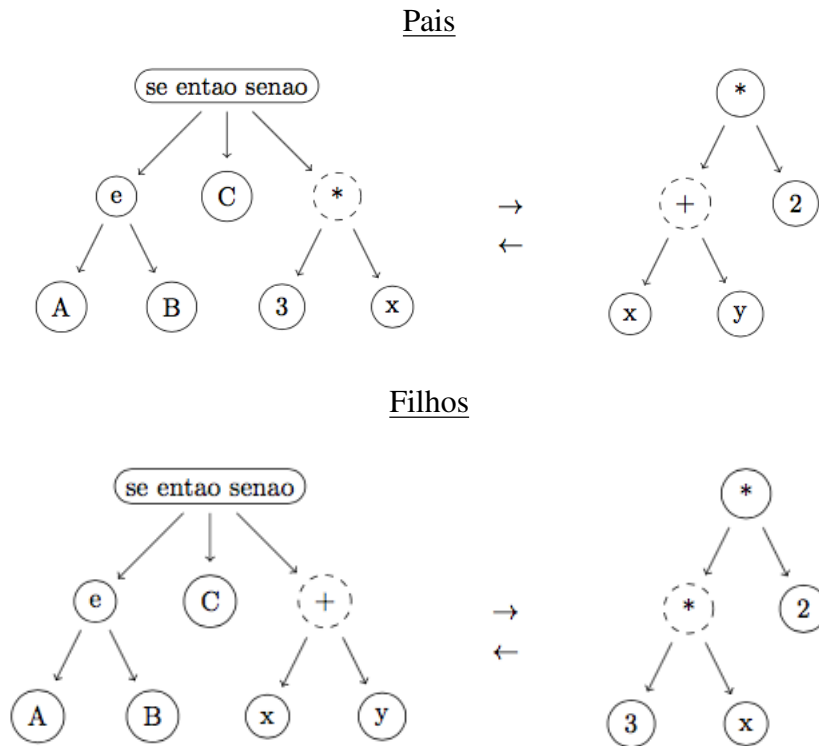
Neste trabalho, foi utilizado o Torneio Binário. Esse método consiste em selecionar aleatoriamente dois indivíduos, comparar suas funções aptidão e escolher o indivíduo de maior valor. Ao fim da seleção, os indivíduos selecionados (mais aptos) são submetidos às fases de “Cruzamento” e/ou “Mutação”.

3.4.5 Operadores Genéticos

Os Operadores Genéticos são usados para gerar uma nova população. Os indivíduos que foram selecionados pelo método de seleção são submetidos aos operadores “Cruzamento” e/ou “Mutação” para gerar uma nova população. O operador “Cruzamento” tem a finalidade de trocar características entre dois indivíduos. Já o operador “Mutação” tem o objetivo de inserir variedade genética na população.

Neste trabalho, foi empregado o “Cruzamento” por ponto de corte, que funciona da seguinte forma: seleciona-se aleatoriamente em cada indivíduo “Pai” uma sub-árvore e, em seguida, trocam-se as sub-árvores dos indivíduos de modo a gerar dois novos indivíduos “Filhos”. A Figura 19 ilustra esse operador.

Figura 19 – Cruzamento por ponto de corte



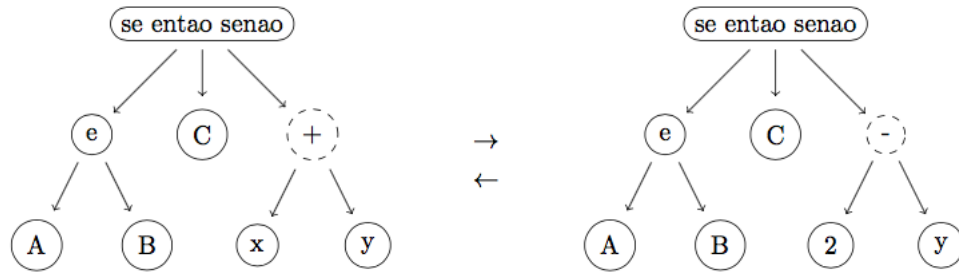
Fonte: (AUGUSTO; BARBOSA, 2013)

A mutação pode se dar de diversas formas. Um dos meios mais comuns consiste em selecionar uma sub-árvore e trocá-la por uma sub-árvore aleatória. A Figura 20 mostra a aplicação aleatória do operador Mutação a um indivíduo. Neste trabalho, foi empregada a mutação *shrink* (KOZA, 1992), que seleciona aleatoriamente um nó da árvore e, em seguida, substitui-o por um nó de \mathbb{T} escolhido de forma aleatória.

3.5 Algoritmos Evolucionários Multiobjetivo

Devido ao fato de ser baseado em população, os Algoritmos Evolucionários (AEs) são bastante adequados para aproximar o conjunto Pareto de um Problema de Otimização Multiobjetivo (POMO). Desde o trabalho de Schaffer (1985), nota-se um grande interesse em utilizar os AEs para resolver os POMO, e esses algoritmos são denominados Algoritmos Evolucionários Multiobjetivo (AEM) (ZHOU et al., 2011).

Figura 20 – Mutação aleatória



Fonte: (AUGUSTO; BARBOSA, 2013)

Os métodos de otimização clássica têm como desvantagens a busca de uma nova solução do conjunto Pareto em cada rodada e a exigência do uso de parâmetros, como por exemplo, λ em P_λ , ou ε_j em ε -Restrito. Esses parâmetros são usados para transformar o POMO em um Problema de Otimização Mono-Objetivo (PO) com o objetivo de encontrar uma única solução, que pode ou não pertencer ao conjunto Pareto. Já os AEs podem ser explorados para enfatizar e preservar a diversidade de soluções não-dominadas de uma população. Desse modo, soluções efetivas podem ser encontradas e mantidas. Depois de algumas gerações, o AE pode conduzir a população a convergir à vizinhança da Fronteira Pareto Ótima com boa diversidade. Os principais AEM são:

1. VEGA: *Vector Enable* GA (SCHAFFER, 1985);
2. MOGA: *Multiobjective* GA (FONSECA; FLEMING, 1995);
3. NPGA: *Niched Pareto* GA (HORN; NAFPLIOTIS; GOLDBERG, 1994);
4. NSGA: *Non-dominated Sorting* GA (SRINIVAS; DEB, 1994);
5. SPEA: *Strength Pareto* EA (ZITZLER et al., 1998);
6. PAES: *Pareto Archived Evolution Strategy* (KNOWLES; CORNE, 1999);
7. PESA: *Pareto Envelop-based Selection Algorithm* (CORNE; KNOWLES; OATES, 2000);
8. PESA-II (CORNE et al., 2001);
9. SPEA2 (ZITZLER et al., 2001);
10. MOIA: *Multiobjective Immune Algorithm* (LUH; CHUEH; LIU, 2003);
11. MOCSA: *Multiobjective Clonal Selection Algorithm* (CAMPELO et al., 2004);
12. IBEA: *Indicator-Based* EA (ZITZLER; KÜNZLI, 2004);

13. NSGA-III (DEB; JAIN, 2014).

Informações sobre esses algoritmos podem ser verificadas nas respectivas referências ou nos livros (COELLO; LAMONT; VELDHUIZEN, 2007) e (DEB; MIETTINEN, 2008).

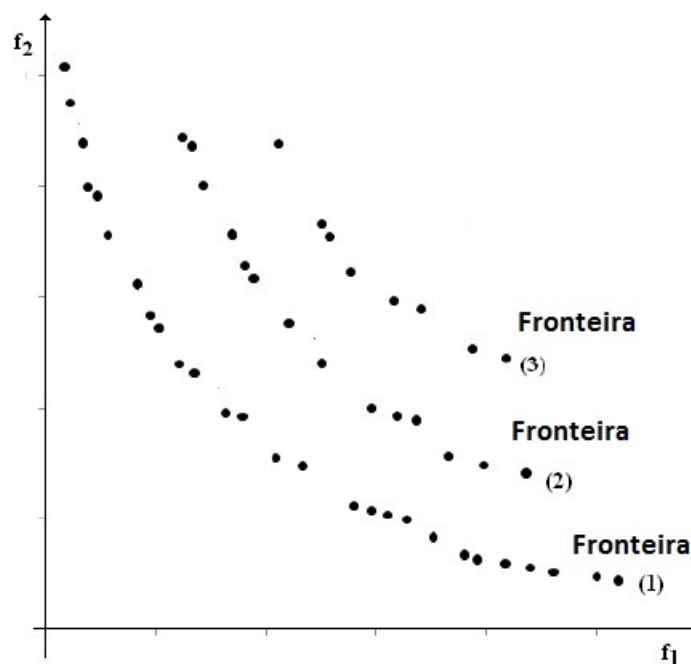
A seguir são apresentados os algoritmos NSGA e NSGA-II (DEB et al., 2002). O NSGA-II é utilizado em um dos sistemas de investimentos propostos.

3.5.1 NSGA

Um dos AEM mais clássicos é o algoritmo *Non-Dominated Sorting Genetic Algorithm* (NSGA) (DEB, 2001). O NSGA utiliza na etapa de seleção do algoritmo genético um esquema de atribuição de *fitness* que prioriza soluções não-dominadas e utiliza uma estratégia de *Sharing* para preservar a diversidade das soluções de cada fronteira não-dominada.

Primeiro, o NSGA particiona a população P em conjuntos $P_i, i = 1, \dots, k$ de soluções não-dominadas. O conjunto $P_1 \subseteq P$ é formado pelas soluções não-dominadas de P . As soluções não-dominadas de $P \setminus P_1$ formam o conjunto P_2 , e assim sucessivamente, até que todas as soluções de P sejam incluídas em algum conjunto. Deb (2001) apresenta um procedimento para realizar essa classificação em $O(MN^3)$. A Figura 21 mostra três Fronteiras Pareto obtidas pelo particionamento de determinada população.

Figura 21 – Exemplo de ordenação pelo conceito de dominância Pareto



Fonte: Elaborado pelo autor

Depois de realizar a classificação, o NSGA atribui *fitness* às soluções de P de modo a priorizar as soluções de P_1 em relação a P_2 , P_2 em relação a P_3 , e assim sucessivamente, até obter P_k , que contém as soluções de menor prioridade. Isso ocorre pois o conjunto P_1 é o conjunto mais próximo à Fronteira Pareto Ótima. Atribuir maior *fitness* às soluções de um melhor conjunto não-dominado assegura uma pressão de seleção em direção à Fronteira Pareto Ótima.

O NSGA mantém a diversidade entre as soluções de cada conjunto. Para isso, o NSGA atribui um menor *fitness* às soluções que tenham um maior número de soluções vizinhas. Desse modo, para cada $P_i, i = 1, \dots, k$, o NSGA aumenta o *fitness* de soluções situadas em regiões pouco povoadas. O NSGA utiliza o modelo de função de compartilhamento para realizar essa atribuição de *fitness*. A função de compartilhamento (3.19) tem o parâmetro d que indica a distância entre duas soluções de P . Se $d = 0$, $f(d) = 1$, o que indica que a solução tem um efeito total sobre si mesma. Por outro lado, se $d \geq \sigma_{sh}$, as duas soluções estão a uma distância de no mínimo σ_{sh} uma da outra, o que produz $f(d) = 0$, que indica que as duas soluções não têm efeito uma sobre a outra. O parâmetro α , quando igual a 1, reduz linearmente o efeito de compartilhamento de 1 a 0.

$$f(d) = \begin{cases} 1 - \left(\frac{d}{\sigma_{sh}}\right)^\alpha, & \text{se } d \leq \sigma_{sh}, \\ 0, & \text{caso contrário.} \end{cases} \quad (3.19)$$

Dado um conjunto $P_i \subseteq P, i \in [1..k]$, para cada par de soluções $\{s, s'\} \in P_i$, calcula-se a distância Euclidiana dada por:

$$d_{s,s'} = \sqrt{\sum_{k=1}^{|P_i|} \left(\frac{x_k^{(s)} - x_k^{(s')}}{x_k^{\max} - x_k^{\min}} \right)^2}. \quad (3.20)$$

O contador de nicho c_s calculado em (3.21) para cada $s \in P_i$ indica uma estimativa da extensão da população de P_i que se encontra ao redor de s .

$$c_s = \sum_{q \in P_i} f(d_{s,q}) \quad (3.21)$$

O algoritmo 2 de atribuição de *Fitness*-NSGA (DEB, 2001) tem como entrada: $\sigma_{sh} = 2$, um positivo $\varepsilon \approx 0$, os conjuntos P_1, P_2, \dots, P_k que são os conjuntos de soluções não-dominadas obtidos pela classificação descrita anteriormente; $N = |P|$; $F^{(s)}$, o *fitness* atribuído à solução s ; e, $F'^{(s)}$, o *fitness* compartilhado atribuído à solução s .

Algorithm 2 Pseudocódigo do algoritmo *Fitness-NSGA*:

```
1:  $F_{min} \leftarrow N + \varepsilon$ .
2: for  $j = 1$  to  $k$  do
3:   for  $s \in P_j$  do
4:      $F^{(s)} \leftarrow F_{min} - \varepsilon$ .
5:     Calcule  $c_s$  por (3.21) utilizando apenas as soluções de  $P_j$ .
6:      $F'^{(s)} \leftarrow \frac{F^{(s)}}{c_s}$ .
7:   end for
8:    $F_{min} = \min\{F'^{(s)} : s \in P_j\}$ .
9: end for
```

Depois de atribuir o *fitness* a todas as soluções de um conjunto $P_i, i = 1, \dots, k-1$, o NSGA atribui o *fitness* às soluções de P_{i+1} um pouco menor que o mínimo *fitness* atribuído às soluções de P_i . Isso garante maior prioridade para os melhores conjuntos de soluções. Desse modo, toda solução de P_1 tem uma chance de sobrevivência maior do que as soluções de P_2 , e assim por diante. Além disso, ao manter a diversidade entre as soluções de determinado conjunto $P_i, i = 1, \dots, k$, o NSGA aumenta a chance de soluções de regiões menos povoadas terem mais cópias na próxima população. Os operadores de *crossover* e mutação são aplicados geralmente a toda população. Mais detalhes sobre o NSGA podem ser encontrados em (DEB, 2001).

3.5.2 NSGA-II

Segundo Deb et al. (2002), o NSGA apresenta problemas como o limite assintótico $O(MN^3)$ para realizar a classificação, a ausência de Elitismo(KOZA, 1992) e a necessidade de especificação dos parâmetros, como σ_{sh} e α . O NSGA-II, proposto por (DEB et al., 2002), é uma extensão do algoritmo original, e apresenta as seguintes vantagens:

1. Implementa elitismo;
2. Utiliza o conceito de distância de multidão, que não depende do ajuste de parâmetros;
3. Apresenta menor complexidade de seleção – $O(MN^2)$.

O NSGA-II usa o *Fast Non-Dominated Sorting*, um método eficiente para a classificação dos indivíduos baseado em não-dominância. Foi proposto também um operador de nicho que tem como objetivo ampliar a diversidade da aproximação do conjunto de Pareto encontrado. Esta técnica foi denominada distância de multidão (*crowding distance*).

Um pseudocódigo do NSGA-II é apresentado no algoritmo 3 a seguir:

Algorithm 3 Pseudocódigo do algoritmo NSGA-II

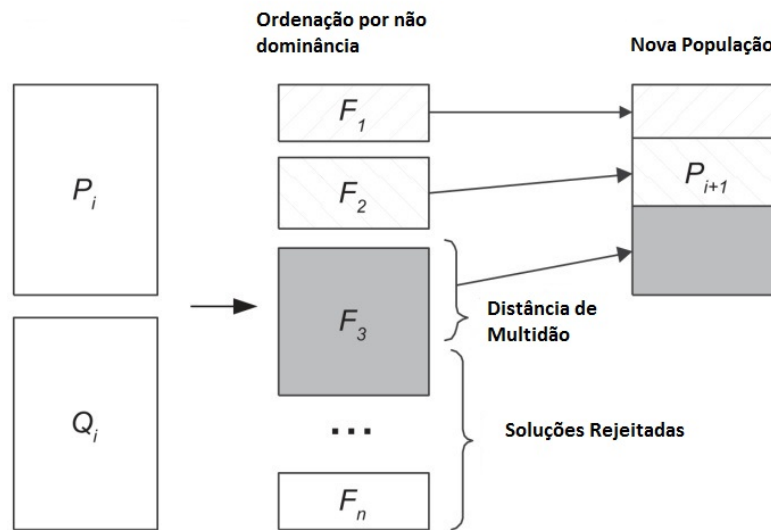
```
1:  $P_1 \leftarrow \text{GeraPopulaoInicial}(N)$ 
2:  $P_1 \leftarrow \text{OrdenaPorNiveisDominancia}(P_1)$ 
3:  $P_1 \leftarrow \text{CalculaDistanciaMultido}(P_1)$ 
4: for  $i = 1$  to  $nger$  do
5:    $Q_i \leftarrow \text{GeraPopulaoFilha}(P_i)$ 
6:    $R_i \leftarrow Q_i \cup P_i$ 
7:    $R_i \leftarrow \text{OrdenaPorNiveisDominancia}(R_i)$ 
8:    $R_i \leftarrow \text{CalculaDistanciaMultido}(R_i)$ 
9:    $P_{i+1} \leftarrow \text{Copia}(R_i, N)$ 
10: end for
```

Onde:

- N : Tamanho fixo para P e Q ;
- P : População pai;
- Q : População filha;
- $nger$: número de gerações;
- GeraPopulaçãoInicial: Gera a população inicial utilizando os parâmetros de um algoritmo evolucionário determinado;
- OrdenaPorNiveisDominância: Utiliza o algoritmo *Fast Non-Dominated Sorting* (DEB et al., 2002) para classificar os indivíduos da população em níveis de dominância. Na figura 21 é apresentado um exemplo;
- CalculaDistânciaMultidão: Utiliza o algoritmo *crowding distance* (DEB et al., 2002), com o objetivo de assegurar a diversidade da aproximação do conjunto de Pareto;
- GeraPopulaçãoFilha: Gera uma população filha aplicando os operadores genéticos (Seleção por Torneio (Nível de Dominância e Distância de Multidão), mutação e cruzamento);
- Copia: Gera uma nova população mantendo os melhores da geração anterior, garantindo o elitismo. Na figura 22 é apresentado um esboço do NSGA-II.

Esse algoritmo inicia com uma população pai P_i de tamanho N . Em seguida, esta é classificada em diferentes níveis de dominância. Para cada indivíduo da população é atribuído um valor de aptidão, igual ao seu nível de dominância (1 para a primeira fronteira, 2 para a seguinte e assim por diante). Os operadores de seleção, cruzamento e mutação são usados sobre a população pai P_i para gerar a população filha Q_i , de mesmo tamanho. O NSGA-II combina as duas populações, P_i e Q_i gerando uma população R_i de tamanho $2N$. O Algoritmo

Figura 22 – Princípio de Seleção do NSGA-II



Fonte: (DEB et al., 2002)

Fast Non-Dominated Sorting é então aplicado em R_i , para formar a próxima geração. Na figura 22 é apresentado o esquema do NSGA-II.

Com objetivo de detalhar o método, são apresentados os pseudocódigos 4 e 5 dos algoritmos *Fast Non-Dominated Sorting* e *crowding-distance-assignment*, conforme descritos em (DEB et al., 2002).

Para finalizar, o NSGA-II utiliza o torneio estocástico clássico (*Stochastic tournament*) como método de seleção. Esse operador utiliza os níveis de dominância e o cálculo de distância de multidão (*Crowding distance*) (ver algoritmo 5). No torneio estocástico, dadas duas soluções x e y , x será vencedora se estiver melhor posicionada em relação às fronteiras das soluções. Se x e y estiverem na mesma fronteira, x é considerada melhor que y se possuir uma maior distância de multidão.

Algorithm 4 Pseudocódigo do *Fast Non-Dominated Sorting*(P)

```
1: for each  $p \in P$  do
2:    $S_p \leftarrow \emptyset$ 
3:    $n_p \leftarrow 0$ 
4:   for each  $q \in P$  do
5:     if  $p < q$  then
6:        $S_p \leftarrow S_p \cup \{q\}$ 
7:     else if  $q < p$  then
8:        $n_p \leftarrow n_p + 1$ 
9:     end if
10:  end for
11:  if  $n_p = 0$  then
12:     $p_{rank} \leftarrow 1$ 
13:     $F_1 \leftarrow F_1 \cup \{p\}$ 
14:  end if
15: end for
16:  $i \leftarrow 1$ 
17: while  $F_i \neq \emptyset$  do
18:    $Q \leftarrow \emptyset$ 
19:   for each  $p \in F_i$  do
20:     for each  $q \in S_p$  do
21:        $n_q \leftarrow n_q + 1$ 
22:       if  $n_q = 0$  then
23:          $q_{rank} \leftarrow i + 1$ 
24:          $Q \leftarrow Q \cup \{q\}$ 
25:       end if
26:     end for
27:   end for
28:    $i \leftarrow i + 1$ 
29:    $F_i \leftarrow Q$ 
30: end while
```

Onde:

- F_i é a i -ésima fronteira do conjunto de soluções como ilustrado nas figuras 21 e 22.

Algorithm 5 Pseudocódigo do *Crowding-distance-assignment* (P)

```
1:  $l \leftarrow |P|$ 
2:  $I[*]_{dist} \leftarrow 0$ ;
3: for  $m = 1$  to  $n_{o\_b\_j}$  do
4:    $I \leftarrow \text{sort}(I, m)$ ;
5:    $I[1]_{dist} \leftarrow I[l]_{dist} \leftarrow \infty$ ;
6:   for  $i = 1$  to  $l - 1$  do
7:      $I[i]_{dist} \leftarrow I[i]_{dist} + \frac{(I[i+1].m - I[i-1].m)}{f_m^{max} - f_m^{min}}$ 
8:   end for
9: end for
```

Onde:

- l é o número de indivíduos na população;
- I é a matriz com os valores de função objetivo;
- $sort(I, m)$ ordena I pelo m -ésimo objetivo;
- $I[i].m$ retorna o valor objetivo m para o indivíduo i ;
- $I[i]_{dist}$ é o valor de *crowding* do indivíduo i .

3.6 Comitê de Decisão

Um Comitê de Decisão constitui-se de um conjunto de classificadores e é utilizado com o objetivo de melhorar o desempenho para predição. A decisão é baseada em treinamento prévio dos classificadores, usando um conjunto de dados de treinamento representativos e para os quais a resposta correta é conhecida. O Comitê de Decisão pondera as decisões individuais de cada classificador de modo a obter um que seja melhor do que cada um individualmente (ROKACH, 2010). Em muitas aplicações, sistemas baseados em comitê têm se mostrado favoráveis quando comparado a sistemas de um único especialista. Alguns classificadores, apesar de terem sido treinados sob as mesmas condições, podem apresentar uma diferente capacidade de generalização ou podem gerar resultados diferentes, particularmente se os dados de teste não forem suficientemente representativos para dados futuros. Nesses casos, a combinação das saídas dos classificadores pela média pode reduzir o risco de seleção de um classificador que é ruim. Em geral, o resultado obtido pela combinação das saídas dos classificadores pode ser ou não melhor do que a saída do melhor classificador do comitê, porém o comitê reduz o risco de uma tomada de decisão muito ruim (POLIKAR, 2006).

Um dos trabalhos iniciais sobre os sistemas comitê de decisão é o trabalho de Dasarathy e Sheela (1979), que particiona o espaço de busca utilizando mais de um classificador. Hansen e Salamon (1990) mostraram que um comitê de redes neurais têm maior capacidade de generalização do que cada rede individual. Schapire (1990) apresenta a técnica *Boosting*, que combina classificadores mais fracos para gerar um classificador mais forte.

Comitês de Decisão são amplamente empregados em inteligência computacional, especialmente em redes neurais (HANSEN; SALAMON, 1990; PERRONE; COOPER, 1992; OPITZ; SHAVLIK, 1996). Nesses casos, vários classificadores, com diferentes complexidades, são utilizados para classificar o mesmo padrão de entrada e seus votos são combinados usando alguma regra específica: a votação majoritária, a média aritmética, a média ponderada, etc.

Em geral, as abordagens utilizadas nos sistemas comitês de decisão diferem pelo procedimento utilizado para gerar classificadores individuais e/ou pela estratégia empregada para

combinar as saídas dos classificadores do comitê. Polikar (2006) apresenta duas formas de combinar os classificadores: seleção e fusão de classificadores. Na seleção, cada classificador é treinado em uma região local do espaço total das soluções. Nesse caso, a combinação é feita a partir da atribuição de pesos mais altos aos classificadores treinados com dados mais próximos dos dados a serem classificados. Um ou mais classificadores locais podem ser escolhidos para a tomada de decisão. Já na fusão, cada classificador é treinado utilizando todo o espaço das soluções. A combinação, nesse caso, é feita via ponderação dos classificadores de modo a obter um classificador mais forte.

Uma das principais características do Comitê de Decisão é obter um conjunto de classificadores com maior capacidade de generalização, ou seja, um comitê que apresenta maior diversidade. Dentre as formas de obter diversidade, estão o uso de diferentes conjuntos de dados de treinamento e/ou diferentes parâmetros. Um comitê de redes neurais *Multi-Layer Perceptron*, por exemplo, pode ser formado por redes com diferentes números de camadas ou diferentes números de nós por camada. A diversidade de um comitê pode ser medida utilizando correlação, q-estatística, medidas de decisões conflitantes, entropia, variância de Kohavi-Wolpert, dentre outros (POLIKAR, 2006).

As principais técnicas utilizadas para gerar um comitê são: *Bagging* (BREIMAN, 1996), *Boosting* (SCHAPIRE, 1990), *AdaBoosting* (FREUND; SCHAPIRE, 1995), dentre outras (POLIKAR, 2006; ROKACH, 2010). Detalhes sobre os algoritmos Bagging, Boosting, AdaBoosting, Mistura de Especialistas (JACOBS et al., 1991), dentre outros, são encontrados em Polikar (2006). A seguir, são apresentados os algoritmos Bagging e Boosting.

Bagging: os classificadores são treinados sob diferentes subconjuntos de dados gerados aleatoriamente a partir de um conjunto de dados disponíveis para treinamento. Em geral, amostras com mais de 75% dos dados são utilizadas para treinamento. A combinação dos classificadores é feita utilizando a estratégia de voto majoritário. Dentre as variações de *Bagging*, estão *Pasting Small Votes* (BREIMAN, 1999) e *Random Forests* (BREIMAN, 2001).

Boosting: Schapire (1990): consiste em gerar um classificador forte, baseando-se em classificadores mais fracos, ou seja, pouco melhores que um classificador aleatório. O algoritmo *Boosting* gera classificadores alterando a amostra de dados. O algoritmo de *Boosting* constrói três classificadores: C_1 , C_2 e C_3 . O classificador C_1 é treinado com um subconjunto aleatório dos dados de treinamento. O C_2 é treinado com um conjunto de dados tal que a metade dos dados é classificada corretamente por C_1 e a outra metade de forma errada por C_1 . O classificador C_3 é treinado com dados sob os quais C_1 e C_2 entram em conflito. Os três classificadores são combinados através de um voto majoritário de modo que C_3 vota quando C_1 e C_2 entram em conflito. Schapire (1990) mostrou que esse classificador, obtido a partir dessa combinação de C_1 , C_2 e C_3 é mais forte. Um classificador mais forte poderia ser obtido empregando-se de forma recursiva o algoritmo de *Boosting*.

Outra abordagem para gerar comitês que tem sido empregada recentemente é a utilização

dos AEs para a construção dos comitês (ABBASS, 2003; CHANDRA; YAO, 2006; LIU; YAO; HIGUCHI, 2000), dentre outros. O trabalho de Abbass (2003), por exemplo, apresenta algoritmos multiobjetivo para gerar comitês de redes neurais artificiais.

Sobre a combinação de classificadores, empregam-se estratégias treináveis ou não. No primeiro caso, os parâmetros do Comitê de Decisão são determinados por algum algoritmo de treinamento. Dentre eles, está o algoritmo Mistura de Especialistas (JACOBS et al., 1991). Já na estratégia sem treinamento, os parâmetros são definidos previamente e utilizados pelo comitê. A votação majoritária ponderada utiliza pesos obtidos sem algoritmo de treinamento.

Sejam dados I classificadores e C classes. Defina as variáveis $d_{ic} \in \{0, 1\}$, $i = 1, \dots, I$, $c = 1, \dots, C$, que indicam se a decisão tomada pelo i -ésimo classificador é a classe c . A votação majoritária possui três versões: todos os classificadores escolhem uma mesma classe, no mínimo 50% mais 1 voto escolhem a classe; ou escolhe-se a classe que recebe o maior número de votos, independentemente de exceder os 50% dos votos.

$$\sum_{i=1}^I d_{ic} = \max \left\{ \sum_{i=1}^I d_{ic} : c = 1, \dots, C \right\} \quad (3.22)$$

Já a votação majoritária ponderada atribui a cada classificador i , $1 \leq i \leq I$, um diferente peso w_i à sua decisão. Neste caso, assume-se que é conhecida previamente uma estimativa do desempenho futuro de cada classificador. Desse modo, os classificadores escolheram a classe c se

$$\sum_{i=1}^I w_i d_{ic} = \max \left\{ \sum_{i=1}^I w_i d_{ic} : c = 1, \dots, C \right\}, \quad (3.23)$$

ou seja, se o total de votos ponderados recebidos por c é maior do que o total de votos recebidos por qualquer outra classe.

4 Sistemas Propostos

Este capítulo tem por finalidade apresentar os sistemas automatizados de investimentos propostos. As duas soluções desenvolvidas têm o objetivo principal de obter lucro no mercado financeiro, identificando, de forma automática, os melhores momentos de compra e de venda de ações em uma janela de tempo determinada. Além de identificar as oportunidades, os sistemas realizam a compra, a venda ou encerram as operações, quando necessário.

Diferentemente de outros tipos de sistemas de investimento, nos quais o objetivo é tentar prever com exatidão o preço da ação no futuro, os sistemas desenvolvidos buscam identificar uma região de compra ou de venda de ações.

Nesse sentido, os sistemas propostos tentam identificar se a série temporal da ação irá subir ou descer nos próximos períodos. Os sistemas são projetados para encontrar soluções que apontem uma compra da ação em tendência de “alta” e uma venda em tendência de “baixa”.

Dadas as técnicas utilizadas, o primeiro sistema a ser apresentado será denominado Sistema Automatizado de Investimento via Seleção Multiobjetivo de Regras (SisSM). Já o segundo, será denominado Sistema Automatizado de Investimento via Programação Genética Multiobjetivo (SisPG).

Os dois sistemas utilizam os indicadores numéricos da Análise Técnica (AT) e possuem os mesmos módulos de negociação. Eles se diferenciam na forma de obter as regras de decisão de compra e de venda.

4.1 Considerações Iniciais

A construção de sistemas automatizados de investimento é uma tarefa antes de tudo de experimentação, de busca, de tentativa e erro. Também é bom ressaltar que a construção destes tipos de sistemas é ainda uma novidade, pois existem poucas disciplinas nos cursos de graduação e pós-graduação no Brasil que abordam este tema, sendo a mesma dificuldade encontrada em conteúdos que exigem uma interdisciplinaridade. No caso de sistema de investimento, são necessários conhecimentos de no mínimo três áreas: finanças, estatística e computação.

Como se já não bastassem os desafios apresentados nesta árdua tarefa de desenvolver sistemas de investimento automatizados, dentro da própria área de Finanças, acontece um controverso debate entre os adeptos da Hipótese do Mercado Eficiente (HME), que não acreditam na possibilidade de obter ganhos no mercado financeiro através da compra e venda de ativos, e os teóricos das Finanças Comportamentais, que acreditam na irracionalidade dos agentes econômicos e nas oportunidades geradas nos movimentos do mercado financeiro.

Nessa tese são apresentados dois sistemas. Porém, ao longo do doutorado foram desenvolvidos e testados pelo menos oito sistemas diferentes. A descrição fria e algorítmica que se segue nas secções seguintes são apenas descrições das técnicas das melhores soluções encontradas. Estão ausentes desta descrição os sistemas que não obtiveram resultados adequados, e também o detalhamento da evolução das soluções propostas até esta se tornar viável.

Ao apresentar estes dois sistemas, o SisSM e o SisPG, procura-se abordar duas técnicas diferentes na sua essência, uma determinista, que busca a otimalidade das regras de compra e venda usando a AT e, a outra, uma abordagem estocástica combinando a AT e a PG. Os dois sistemas fazem uso da otimização biobjetivo. A ordem de apresentação dos sistemas nesta tese se refere ao grau de dificuldade empenhado para construí-los e na qualidade dos resultados apresentados. Pode-se considerar que o SisPG é uma evolução do SisSM, tanto do ponto de vista das técnicas utilizadas, quanto da estabilidade dos resultados.

Nas secções seguintes são detalhados cada um dos sistemas propostos e são apresentados exemplos do seu funcionamento.

4.2 Sistema Automatizado de Investimento via Seleção Multiobjetivo de Regras (SisSM)

O SisSM recebe uma série financeira temporal de um determinado ativo, contendo as seguintes informações: (i) o preço de abertura, (ii) o preço de fechamento, (iii) o preço mínimo, (iv) o preço máximo e o (v) volume negociado. Esses dados são usados como entrada para o cálculo de indicadores da AT, que, por sua vez, serão a base para a formação das regras de compra e regras de venda dos ativos.

O SisSM trabalha em três janelas móveis de tempo, sendo que, na primeira janela, é realizada a seleção de regras fase I; na segunda janela é realizada a seleção de regras fase II e a montagem do comitê de decisão; na terceira e última janela é feita a negociação das ações.

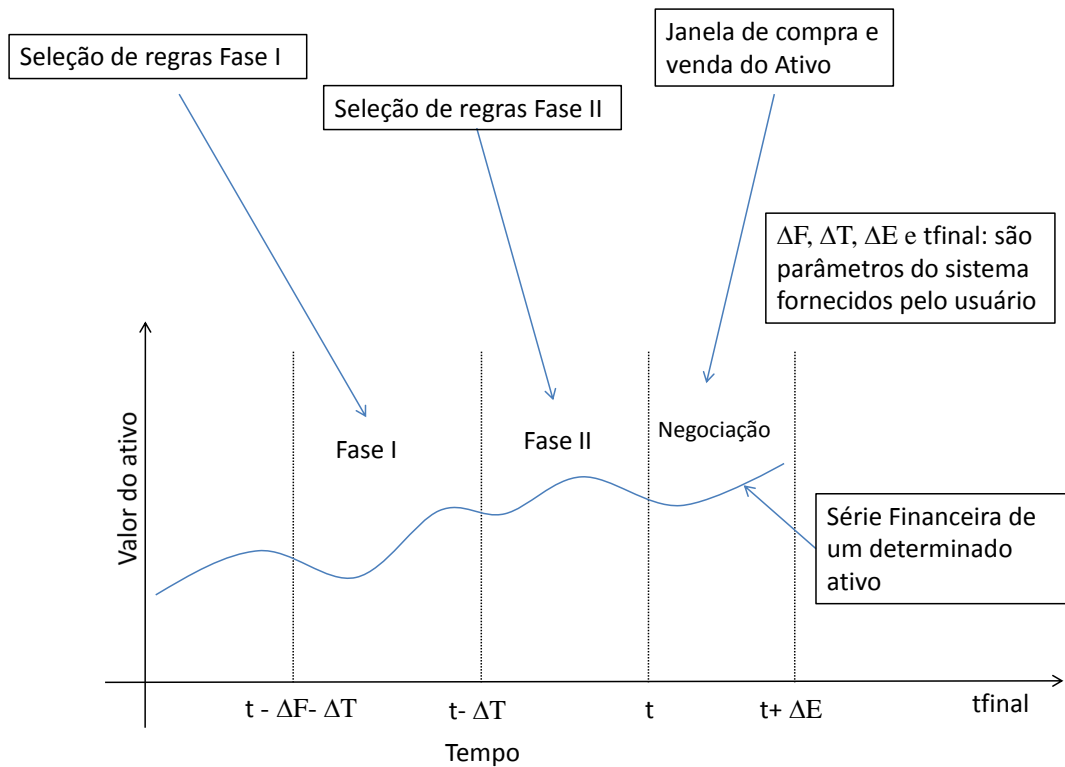
Cada solução no SisSM é composta pela combinação de uma regra de compra e uma regra de venda. Foram formuladas 34 regras de compra e 34 regras de venda, possibilitando a formação de 1.156 soluções diferentes.

O sistema funciona da seguinte forma: na primeira janela móvel, seleção de regras fase I, são selecionadas as N_a melhores combinações. Entende-se por melhor combinação aquela solução composta por uma regra de compra e uma regra de venda que obteve o maior retorno financeiro. Na segunda janela móvel, seleção de regras fase II, são testadas as N_a soluções da fase I e selecionadas as N_b melhores soluções. Todas soluções geradas da fase II são submetidas à janela de negociação. A cada iteração, o sistema contabiliza o retorno financeiro de todas as soluções refinadas da fase II, que é acumulado ao longo das iterações.

Para finalizar, o Comitê de Decisão da primeira iteração é formado pelas N_b melhores

soluções da fase II. A partir da segunda iteração, o comitê é formado com base no desempenho das soluções testadas na fase de negociação já contabilizadas. Esse desempenho é aferido a partir de duas funções objetivo que são apresentadas nas subseções seguintes. N_a e N_b são parâmetros definidos pelo usuário. Na figura 23, é apresentado o primeiro esboço das fases do SisSM.

Figura 23 – Fases do SisSM

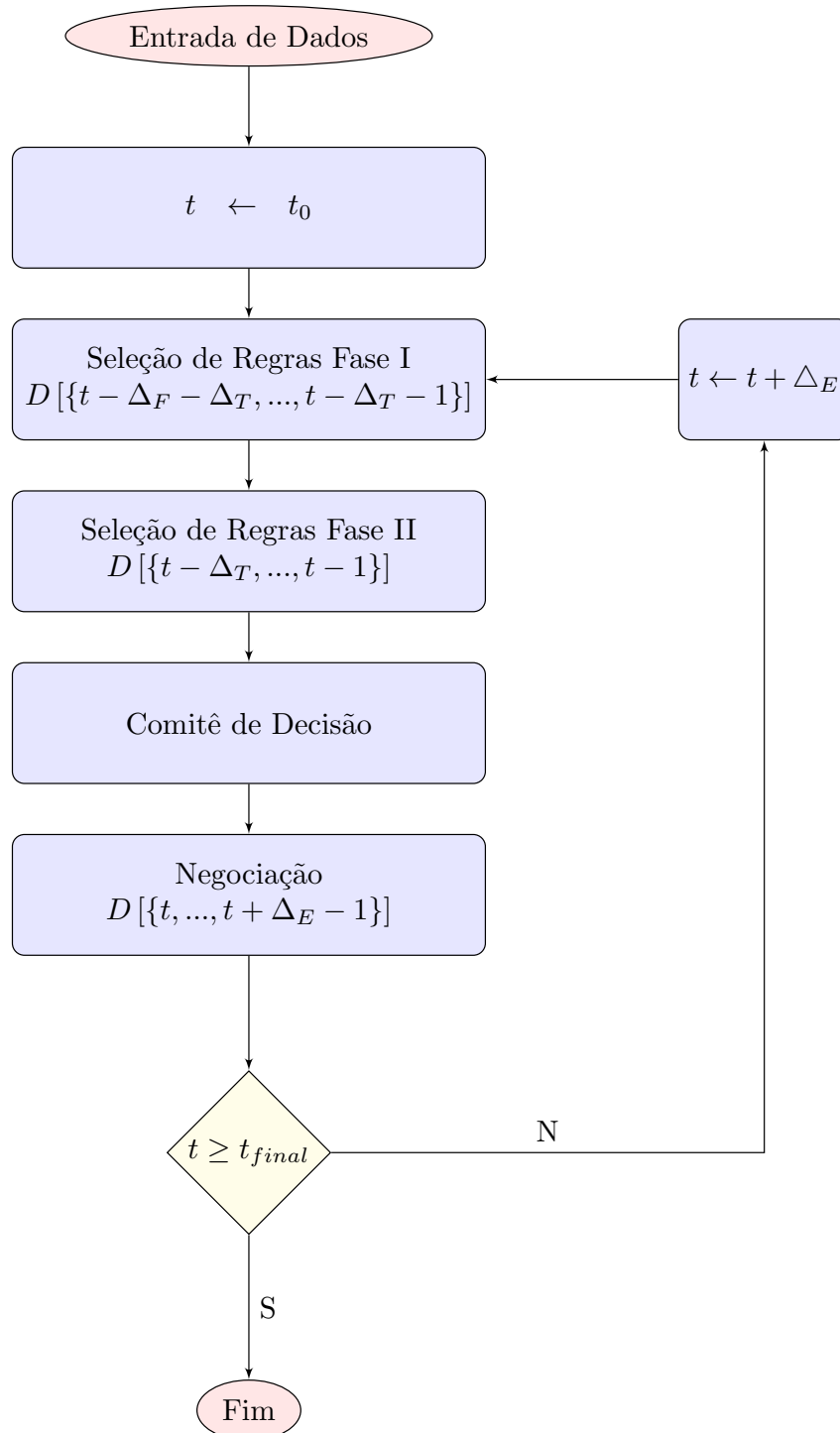


Fonte: Elaborada pelo autor

4.2.1 Visão Geral do Sistema do SisSM

O esquema geral do SisSM é apresentado na figura 24. O sistema trabalha com janelas móveis dentro de um laço de repetição, em que a condição de parada é quando a data final (t_{final}) é atingida. Três janelas de tempo são usadas em cada iteração: seleção de regras fase I $\{t - \Delta F - \Delta T, \dots, t - \Delta T - 1\}$ (dados passados), seleção de regras fase II $\{t - \Delta T, \dots, t - 1\}$ (dados passados) e a janela de investimentos $\{t, \dots, t + \Delta E - 1\}$ (dados futuros). O sistema proposto pode ser resumido em quatro operações principais, que são realizadas em cada iteração:

Figura 24 – Esquema Geral do SisSM. Neste esquema: $D[\{t_I, \dots, t_F\}]$ é a série financeira inicial t_I e final t_F ; $\Delta_F > 0$, $\Delta_T > 0$, e $\Delta_E > 0$ tamanho das janelas de seleção de características, treinamento e negociação, respectivamente



Fonte: Elaborada pelo autor

1. Seleção de Regras Fase I $D[\{t - \Delta_F - \Delta_T, \dots, t - \Delta_T - 1\}]$: seleciona as N_a melhores combinações em um universo de 1156 soluções diferentes, como apresentado nas tabelas

- 4 e 5. A melhor combinação é aquela solução composta por uma regra de compra e uma regra de venda que obtiveram o maior retorno financeiro em uma dada janela de tempo.
2. Seleção de Regras Fase II $D[\{t - \Delta_T, \dots, t - 1\}]$: seleciona as N_b melhores combinações de regras de compra e venda, em um universo de N_a soluções obtidas da fase I.
 3. Comitê de Decisão: na primeira iteração, o Comitê de Decisão é formado pelas N_b melhores soluções obtidas da fase II. A partir da segunda iteração, o comitê é formado com base no desempenho das soluções testadas na fase de negociação.
 4. Negociação $D[\{t, \dots, t + \Delta_E - 1\}]$: o comitê de decisão é usado para identificar se uma das três operações serão realizadas no final de cada dia na janela de avaliação: (i) comprar; (ii) vender, ou; (iii) manter.

Nas próximas subseções são apresentados os algoritmos do SisSM, juntamente com cada uma das etapas descritas. $N_a, N_b, t_0, t_{final}, \Delta_T, \Delta_F, \Delta_E$ são parâmetros definidos pelo usuário.

4.2.2 Algoritmo geral do SisSM

O SisSM foi elaborado com o intuito de permitir que as regras mais bem sucedidas, ao longo das várias iterações, fossem selecionadas para participar das decisões de compra e de venda. As melhores regras são selecionadas dinamicamente, por meio do desempenho de duas funções objetivo descritas nas equações 4.1 e 4.2.

A cada iteração é feita uma nova seleção de regras para participar do comitê de decisão. O algoritmo geral do SisSM é apresentado a seguir, no pseudocódigo 6:

Algorithm 6 Pseudocódigo do SisSM

```
 $t \leftarrow t_0;$ 
 $RulesEnsemble \leftarrow \emptyset;$ 
 $RulesCandidate \leftarrow \emptyset;$ 
 $VarTrading.Status \leftarrow 0;$ 
 $VarTrading.ProfitLoss \leftarrow 0;$ 
 $VarTrading.Cash \leftarrow InitialCapital;$ 
 $VarTrading.Share \leftarrow 0;$ 
 $(SetRulesBuy, SetRulesSell) \leftarrow \mathbf{GenerateAllRules};$ 
while  $t < t_{final}$  do
   $SlidingWindowI \leftarrow \mathbf{SelectData}(FinancialSeries, t - \Delta_F - \Delta_T, t - \Delta_T - 1);$ 
   $SlidingWindowII \leftarrow \mathbf{SelectData}(FinancialSeries, t - \Delta_T, t - 1);$ 
   $TradingWindow \leftarrow \mathbf{SelectData}(FinancialSeries, t, t + \Delta_E - 1);$ 
   $RulesI \leftarrow \mathbf{RulesSelectionI}(SlidingWindowI, SetRulesBuy, SetRulesSell, N_a);$ 
   $RulesII \leftarrow \mathbf{RulesSelectionII}(SlidingWindowII, RulesI, N_b);$ 
  if  $t = t_0$  then
     $RulesEnsemble \leftarrow RulesII;$ 
  else
     $RulesEnsemble \leftarrow \mathbf{FastNonDominatedSorting}(RulesCandidate, F_{a1}, F_{a2});$ 
  end if
   $VarTrading \leftarrow \mathbf{TradingModule}(RulesEnsemble, TradingWindow, VarTrading);$ 
   $RulesCandidate \leftarrow RulesCandidate \cup RulesII;$ 
   $RulesCandidate \leftarrow \mathbf{RulesTradingModule}(RulesCandidate, TradingWindow);$ 
   $t \leftarrow t + \Delta_E;$ 
end while
 $ProfitLoss \leftarrow VarTrading.ProfitLoss;$ 
```

Onde:

- $t_0, t_{final}, \Delta_F, \Delta_T, \Delta_E, N_a$: são parâmetros fornecidos pelo usuário;
- *VarTrading*: é uma variável estruturada com quatro campos distintos, de fundamental importância para o controle do SisSM. O campo **VarTrading.Status** pode assumir três valores que indicam a situação do sistema no *trading*: o valor 0 significa que o sistema está fora do *trading*, o valor 1 significa que o sistema está na situação “comprado” e o valor 2 significa que o sistema está na situação “vendido”. O campo **VarTrading.Cash** registra o quanto de valor o sistema tem em caixa para negociar a compra ou a venda de ações. O campo **VarTrading.ProfitLoss** armazena os lucros e prejuízos que o sistema obtém ao longo das N janelas de negociação. O campo **VarTrading.Share** registra a quantidade de ações disponíveis. É importante destacar que esse campo pode ter valor negativo, considerando que é permitido vender ações a descoberto (ELDER, 1993);
- *RulesEnsemble*: esta variável armazena as soluções que farão parte do Comitê de Decisão;
- *RulesCandidate*: esta variável armazena as soluções candidatas a participarem do Comitê de Decisão;

- **GenerateAllRules**: este procedimento gera todas as soluções candidatas. Ver as tabelas 4 e 5;
- **RulesI**: armazena as N_a regras obtidas a partir da primeira janela móvel;
- **RulesII**: armazena as N_b regras obtidas a partir da segunda janela móvel;
- **RulesSelectionI**: este algoritmo descrito no pseudocódigo 7, seleciona as N_a melhores soluções a partir da primeira janela móvel;
- **RulesSelectionII**: este algoritmo descrito no pseudocódigo 8, seleciona as N_b melhores soluções a partir da segunda janela móvel;
- **FastNonDominatedSorting**: este algoritmo descrito no pseudocódigo 4 foi proposto por (DEB et al., 2002) e descrito em detalhes no capítulo 3, seção 3.5;
- **TradingModule**: este algoritmo é responsável pela negociação das ações; é o mesmo para os dois sistemas, o SisPG e o SisSM. Sua descrição é detalhada no pseudocódigo 10;
- **RulesTradingModule**: este módulo calcula o retorno financeiro de cada solução candidata em uma dada série financeira. Uma solução candidata é composta por uma regra de compra e uma regra de venda e passou pelos filtros de seleção de regras fase I e fase II. Portanto, existe a probabilidade dessa solução fazer parte do Comitê de Decisão.

4.2.3 Seleção de Regras Fase I do SisSM

Nos dois sistemas desenvolvidos, o SisSM e o SisPG, foram utilizados 12 indicadores técnicos, apresentados na tabela 3. Estes indicadores foram escolhidos por conterem os quatro grupos principais da AT computadorizada: os seguidores de tendência, os osciladores, os sistemas de bandas e os identificadores de divergência. Também foi considerada a facilidade de se calcular cada indicador. Todos eles podem ser obtidos a partir de cálculos feitos sob os valores das séries temporais, como o valor da ação na abertura do mercado, o valor de fechamento, a máxima do dia, a mínima do dia e o volume financeiro negociado.

As regras de compra e de venda apresentadas nas tabelas 4 e 5 são todas as regras válidas de compra e de venda utilizadas pelo SisSM. Ao todo, são contabilizadas 34 regras de compra e 34 regras de venda. Uma solução candidata é formada por uma regra de compra e uma regra de venda. Essas regras devem ser lidas como o exemplo a seguir:

- **Compra** :: $IFR(i) < 15$: Se o Índice de Força Relativa (IFR) é menor que 15, então compre. O ativo está sobrevendido.
- **Venda** :: $IFR(i) > 95$: Se o IFR é maior que 95, então venda. O ativo está sobrecomprado.

Tabela 3 – Indicadores Numéricos da AT

Indicador	Nome	Categoria
MMS	Média Móvel Simples	Seguidores de Tendência
MME	Média Móvel Exponencial	Seguidores de Tendência
OC	Oscilador de Chaikin	Osciladores
VC	Volatilidade de Chaikin	Osciladores
WR	Williams R	Osciladores
VROC	Volume Taxa de Variação	Osciladores
PROC	Preço Taxa de Variação	Osciladores
BB	Bandas de Bolling	Sistemas de Bandas
AD	Acumulação e Distribuição	Identificadores de Divergência
OBV	Saldo de Volume	Identificadores de Divergência
IFR	Índice de Força Relativa	Identificadores de Divergência
E	Estocástico	Identificadores de Divergência

Fonte: Elaborada pelo autor

As regras descritas nas tabelas 4 e 5 foram inspiradas nas regras apresentadas em livros de AT, tais como (MURPHY, 1999; ACHELIS, 2001; ELDER, 1993).

A primeira fase da seleção de regras consiste em buscar, em um conjunto de 1.156 combinações, as N_a soluções mais rentáveis testadas em uma determinada janela de tempo. Nessa fase, é feito o primeiro filtro de possíveis soluções que poderão participar do Comitê de Decisão. O algoritmo da seleção de regras fase I é apresentado a seguir, no pseudocódigo 7.

Algorithm 7 Pseudocódigo da Seleção de Regras Fase I

```

function RULESELECTIONI(SlidingWindowI, SetRuleBuy, SetRuleSell,  $N_a$ )
   $L \leftarrow \text{length}(\text{SetRuleBuy})$ ;
   $K \leftarrow \text{length}(\text{SetRuleSell})$ ;
  for  $n = 1$  to  $L$  do
    for  $m = 1$  to  $K$  do
       $\text{SetAllRule} \leftarrow \text{Account}(\text{SetRuleBuy}(n), \text{SetRuleSell}(m), \text{SlidingWindowI})$ ;
    end for
    end for
     $\text{SetAllRule} \leftarrow \text{Sort}(\text{SetAllRule}, 'descendent')$ ;
     $\text{SubSetRule} \leftarrow \text{SelectRule}(\text{SetAllRule}, 1 : N_a)$ ;
    return ( $\text{SubSetRule}$ );
end function

```

Onde:

- *SlidingWindowI*: é a série financeira de onde se pretende selecionar as melhores regras de negociação;
- *SetRulesBuy*: conjunto de regras de compra apresentadas na tabela 4;
- *SetRuleSell*: conjunto de regras de venda apresentadas na tabela 5;

Tabela 4 – Regras lógicas de compra baseadas em indicadores numéricos da AT do SisSM

Ind	Cod	Regras de Compra
1	RC01	$\text{And}(\text{PrecoAtivo}(i-1) < \text{BBI}_{20}(i-1), \text{PrecoAtivo}(i) > \text{BBI}_{20}(i))$
2	RC02	$\text{And}(\text{PrecoAtivo}(i-1) < \text{BBI}_{14}(i-1), \text{PrecoAtivo}(i) > \text{BBI}_{14}(i))$
3	RC03	$\text{And}(\text{PrecoAtivo}(i-1) < \text{BBI}_{24}(i-1), \text{PrecoAtivo}(i) > \text{BBI}_{24}(i))$
4	RC04	$(\text{IFR}_2(i, 1) < 10)$
5	RC05	$(\text{IFR}_{14}(i, 1) < 30)$
6	RC06	$(\text{IFR}_{14}(i, 1) < 20)$
7	RC07	$(\text{WR}_{14}(i) < -90)$
8	RC08	$(\text{WR}_{21}(i) < -90)$
9	RC09	$(E_{14} < 20)$
10	RC10	$(E_5 < 20)$
11	RC11	$(\text{IFR}_2(i, 1) < 15)$
12	RC12	$(\text{IFR}_2(i, 1) < 10)$
13	RC13	$(\text{VROC}_{14} > 120)$
14	RC14	$(\text{VROC}_{14} > 150)$
15	RC15	$\text{And}(\text{AD}(i) > \text{AD}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-1))$
16	RC16	$\text{And}(\text{AD}(i) > \text{AD}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-2))$
17	RC17	$\text{And}(\text{AD}(i) > \text{AD}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-3))$
18	RC18	$\text{And}(\text{OBV}(i) > \text{OBV}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-1))$
19	RC19	$\text{And}(\text{OBV}(i) > \text{OBV}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-2))$
20	RC20	$\text{And}(\text{OBV}(i) > \text{OBV}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-3))$
21	RC21	$\text{And}(\text{OBV}(i) > \text{OBV}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-4))$
22	RC22	$\text{And}(\text{OC}(i) > 0, \text{OC}(i-1) < 0)$
23	RC23	$\text{And}(\text{OC}(i) > 0, \text{OC}(i-2) < 0)$
24	RC24	$\text{And}(\text{OC}(i) > 0, \text{OC}(i-3) < 0)$
25	RC25	$(\text{VC}(i) < 0)$
26	RC26	$(\text{PROC}(i) < -10)$
27	RC27	$(\text{PrecoAtivo}(i) < \text{MME}_{21})$
28	RC28	$(\text{PrecoAtivo}(i-1) < \text{MME}_{21})$
29	RC29	$(\text{PrecoAtivo}(i-2) < \text{MME}_{21})$
30	RC30	$(\text{PrecoAtivo}(i-3) < \text{MME}_{21})$
31	RC31	$(\text{PrecoAtivo}(i) < \text{MME}_5)$
32	RC32	$(\text{PrecoAtivo}(i-1) < \text{MME}_6)$
33	RC33	$(\text{PrecoAtivo}(i-2) < \text{MME}_7)$
34	RC34	$(\text{PrecoAtivo}(i-3) < \text{MME}_8)$

Fonte: Elaborada pelo autor

Tabela 5 – Regras lógicas de venda baseadas em indicadores numéricos da AT do SisSM

Ind	Cod	Regras de Venda
1	RV01	$\text{And}(\text{PrecoAtivo}(i-1) < BBS_{20}(i-1), \text{PrecoAtivo}(i) > BBS_{20}(i))$
2	RV02	$\text{And}(\text{PrecoAtivo}(i-1) < BBS_{14}(i-1), \text{PrecoAtivo}(i) > BBS_{14}(i))$
3	RV03	$\text{And}(\text{PrecoAtivo}(i-1) < BBS_{24}(i-1), \text{PrecoAtivo}(i) > BBS_{24}(i))$
4	RV04	$(IFR_2(i, 1) > 95)$
5	RV05	$(IFR_{14}(i, 1) > 70)$
6	RV06	$(IFR_{14}(i, 1) > 80)$
7	RV07	$(WR_{14}(i) > -10)$
8	RV08	$(WR_{21}(i) > -10)$
9	RV09	$(E_{14} > 90)$
10	RV10	$(E_5 > 90)$
11	RV11	$(IFR_2(i, 1) > 95)$
12	RV12	$(IFR_2(i, 1) > 97)$
13	RV13	$(VROC_{14} > 120)$
14	RV14	$(VROC_{14} > 150)$
15	RV15	$\text{And}(AD(i) < AD(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-1))$
16	RV16	$\text{And}(AD(i) < AD(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-2))$
17	RV17	$\text{And}(AD(i) < AD(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-3))$
18	RV18	$\text{And}(OBV(i) < OBV(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-1))$
19	RV19	$\text{And}(OBV(i) < OBV(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-2))$
20	RV20	$\text{And}(OBV(i) < OBV(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-3))$
21	RV21	$\text{And}(OBV(i) < OBV(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-4))$
22	RV22	$\text{And}(OC(i) < 0, OC(i-1) > 0)$
23	RV23	$\text{And}(OC(i) < 0, OC(i-2) > 0)$
24	RV24	$\text{And}(OC(i) < 0, OC(i-3) > 0)$
25	RV25	$(VC(i) > 100)$
26	RV26	$(PROC(i) > 10)$
27	RV27	$(\text{PrecoAtivo}(i) > MME_{21})$
28	RV28	$(\text{PrecoAtivo}(i-1) > MME_{21})$
29	RV29	$(\text{PrecoAtivo}(i-2) > MME_{21})$
30	RV30	$(\text{PrecoAtivo}(i-3) > MME_{21})$
31	RV31	$(\text{PrecoAtivo}(i) > MME_5)$
32	RV32	$(\text{PrecoAtivo}(i-1) > MME_6)$
33	RV33	$(\text{PrecoAtivo}(i-2) > MME_7)$
34	RV34	$(\text{PrecoAtivo}(i-3) > MME_8)$

Fonte: Elaborada pelo autor

- N_a : número de regras a serem selecionadas; parâmetro definido pelo usuário.

4.2.4 Seleção de Regras Fase II do SisSM

A segunda fase da seleção de regras consiste em buscar, em um conjunto de N_a soluções provenientes da fase seleção de regras I, aquelas que são mais rentáveis, testadas em uma determinada janela de tempo. Nesta fase, é feito o segundo filtro de possíveis soluções que poderão participar do Comitê de Decisão. É importante observar que o segundo filtro é feito em uma janela de tempo distinta da fase anterior.

O algoritmo da seleção de regras fase II é apresentado a seguir, no pseudocódigo 8.

Algorithm 8 Pseudocódigo da Seleção de Regras Fase II

```

function RULESELECTIONII(SlidingWindowII, RulesI,  $N_b$ )
   $L \leftarrow \text{length}(\text{RulesI})$ ;
  for  $n = 1$  to  $L$  do
     $\text{SubSetRule} \leftarrow \text{Account}(\text{RulesI}(n), \text{SlidingWindowII})$ ;
  end for
   $\text{SetRule} \leftarrow \text{Sort}(\text{SubSetRule}, 'descendent')$ ;
   $\text{SetRule} \leftarrow \text{SelectRule}(\text{SetRule}, 1 : N_b)$ ;
  return ( $\text{SetRule}$ );
end function

```

Onde:

- *SlidingWindowII*: é a série financeira de onde se pretende selecionar as melhores regras de negociação;
- *SetRule*: conjunto de regras selecionadas;
- N_b : número de regras a serem selecionadas, parâmetro definido pelo usuário.

4.2.5 Funcionamento e Formação do Comitê de Decisão do SisSM

A composição do Comitê de Decisão é feita sobre as regras previamente selecionadas. Essa seleção funciona da seguinte forma: na primeira iteração, as N_b melhores soluções provenientes da fase de seleção de regras II são destinadas ao Comitê de Decisão. A partir da segunda iteração, o comitê é constituído com base no desempenho das soluções já contabilizadas na fase de negociação. O desempenho das soluções já testadas é verificado por meio de duas funções objetivo, F_{a1} e F_{a2} , como apresentadas a seguir, nas equações 4.1 e 4.2 respectivamente.

Na segunda iteração, a seleção de regras para o Comitê de Decisão é tratada como um problema de maximização biobjetivo, sendo utilizado o operador *Fast Non Dominated Sorting* (DEB et al., 2002) para escolher as soluções não dominadas.

A função objetivo F_{a1} contabiliza, ao longo de cada iteração, os lucros e prejuízos de cada solução. A proposta é “premiar” as soluções que tenham um histórico de melhor retorno financeiro. Nessa equação 4.1, $K_1(p)$ é o retorno financeiro de uma solução avaliada em um período p . A cada iteração, p é incrementado de Δ_E .

$$F_{a1} = \sum_{p=t_0}^{t_f} K_1(p) \quad (4.1)$$

A função objetivo F_{a2} soma +1, ao longo de cada iteração, se houver lucro, e subtrai -1, se houver prejuízo. O objetivo dessa segunda função é “premiar” as soluções que apresentem maior constância em retorno financeiros positivos, independentemente, do seu valor absoluto. Essa função pode ser entendida também como um moderador, diminuindo as distorções dos retornos. Nessa equação 4.2, $K_2(p)$ é contabilizado em um período p , da seguinte forma:

1. se o retorno financeiro de uma solução avaliada em um período p for maior que 0 , $K_2(p)$ retorna 1;
2. se o retorno financeiro de uma solução avaliada em um período p for menor que 0, $K_2(p)$ retorna -1;
3. A cada iteração p é incrementado de Δ_E .

$$F_{a2} = \sum_{p=t_0}^{t_f} K_2(p) \quad (4.2)$$

Já a decisão de comprar ou de vender uma ação é feita por um sistema de votação da maioria. Cada regra de compra e de venda de cada solução é avaliada, dia a dia. Se a maioria indicar uma compra, é retornado um sinal de compra. Se a maioria indicar uma venda, é retornado um sinal de venda. No algoritmo 9 é detalhado o funcionamento do Comitê de Decisão.

Algorithm 9 Pseudocódigo do Comitê de Decisão

```
function ENSEMBLE(SetRules, CurrentDay)
  ExpBuy  $\leftarrow$  false;
  ExpSell  $\leftarrow$  false;
  cBuy  $\leftarrow$  0;
  cSell  $\leftarrow$  0;
  L  $\leftarrow$  length(SetRules);
  for n = 1 to L do
    SetupBuy  $\leftarrow$  Eval(SetRules(n).Buy, CurrentDay);
    SetupSell  $\leftarrow$  Eval(SetRules(n).Sell, CurrentDay);

    if XOR(SetupBuy, SetupSell) then

      if SetupBuy then
        cBuy  $\leftarrow$  cBuy + 1;
      else
        cSell  $\leftarrow$  cSell + 1;
      end if
    end if
  end for

  if (cBuy  $\neq$  cSell) then
    if cBuy > cSell then
      ExpBuy  $\leftarrow$  true;
      ExpSell  $\leftarrow$  false;
    else
      ExpBuy  $\leftarrow$  false;
      ExpSell  $\leftarrow$  true;
    end if
  end if
  return (ExpBuy, ExpSell);
end function
```

Onde:

- *SetRules*: conjunto de soluções recebidas da fase de Seleção II;
- *CurrentDay*: data corrente de negociação; as decisões são tomadas dia a dia.

4.2.6 Sistema de negociação do SisSM

O último componente do SisSM é o módulo de negociação. Esse é responsável pela execução das ordens de compra ou de venda. Esse módulo foi construído com base em seis características principais:

- as decisões são feitas por meio de um Comitê de Decisão;

- é possível fazer o aluguel de ações, se necessário;
- alavanca o capital em uma vez o valor do saldo disponível;
- compra e vende no fechamento de mercado;
- as negociações são interrompidas quando se atinge uma perda máxima (*MaxLoss*) ou quando se atinge um ganho máximo (*MaxGain*);
- as posições são fechadas no final da janela de negociação.

Quanto ao aluguel de ações, quando o sistema indica uma venda, é permitido que ela seja negociada, sem necessariamente possuir uma ação. Esse tipo de operação é uma prática do mercado de ações, denominada venda a descoberto (MURPHY, 1999; ELDER, 1993).

Outra prática comum no mercado de ações é alavancar (MURPHY, 1999; ELDER, 1993) nas operações de compra ou de venda. Em finanças, alavancagem é o termo geral para qualquer técnica aplicada para multiplicar a rentabilidade, por meio de endividamento. O incremento proporcionado por meio da alavancagem também aumenta os riscos da operação.

Outra característica do sistema de negociação é o momento da compra ou da venda das ações. O sistema trabalha com *candles* diários e toda compra ou venda é feita no preço de fechamento do dia. Esta decisão foi tomada devido à alta liquidez da bolsa de valores no fechamento do dia (ELDER, 1993).

O módulo de negociação é aplicado em cada dia da janela de avaliação. Após cada execução, o sistema deve estar em um dos quatro estados: (i) fora do mercado (início); (ii) comprado; (iii) vendido, ou; (iv) fora do mercado (fim). As ações executadas variam de acordo com o estado atual. O diagrama de transição do módulo de negociação proposto é apresentado na figura 25.

A seguir é apresentada uma descrição do procedimento e o algoritmo proposto é detalhado no pseudocódigo 10:

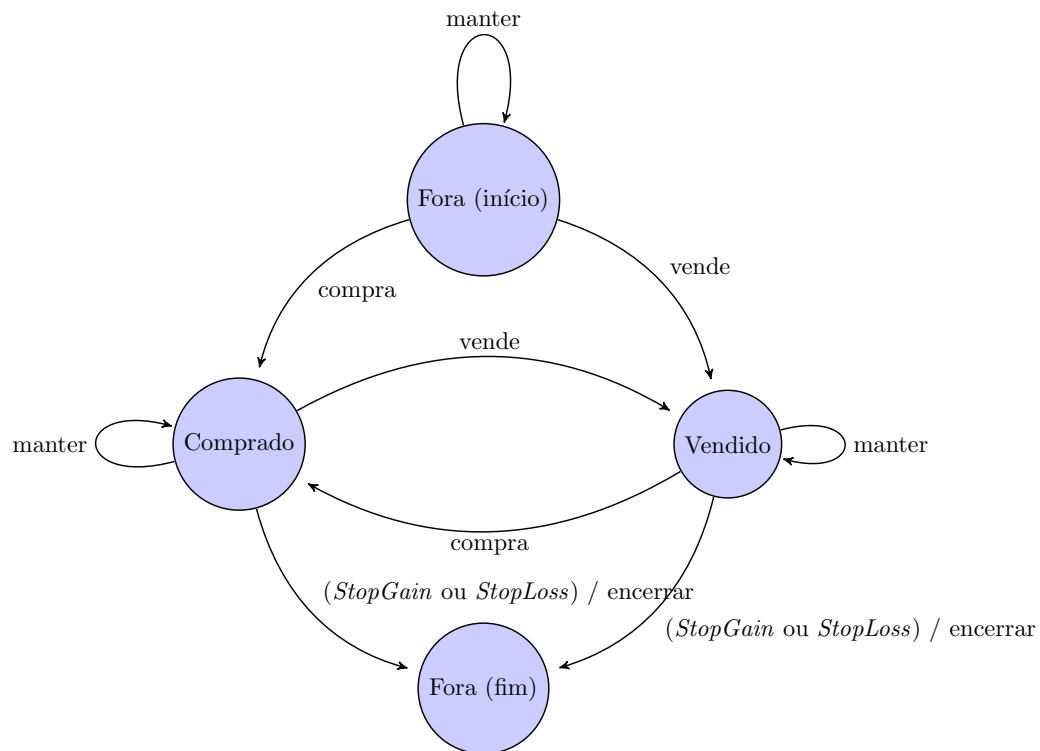
Fora do mercado (início): Este é o estado no primeiro dia da janela de avaliação. O estado do sistema no dia seguinte será definido pela ação indicada pelo Comitê de Decisão:

- Manter: o sistema permanecerá no estado (início) fora do mercado.
- Comprar: será utilizado o dinheiro disponível para comprar ações e passará para o estado comprado.
- Vender: será efetuado o aluguel e a venda de ações e passará para o estado vendido.

Comprado: nesse estado, o Comitê de Decisão indicará uma de duas transições possíveis:

- Manter no estado comprado: o sistema permanecerá no estado comprado.

Figura 25 – Diagrama de transição do módulo de negociação. Na figura, Fora (início) e Fora (fim) representam o estado (inicial) e o estado (final) do diagrama



Fonte: Elaborada pelo autor

- Vender: vende todas as ações disponíveis e passa para o estado vendido.

Além disso, o sistema pode passar para o estado Fora do mercado (final) se uma das seguintes condições se tornar válida: (i) *stop loss* é atingido; (ii) *stop gain* é atingido, ou (iii) é o último dia da janela de avaliação.

Vendido: neste estado, o Comitê de Decisão indicará uma de duas transições possíveis:

- Manter no estado vendido: o sistema permanecerá no estado vendido.
- Comprar: será utilizado o dinheiro disponível para comprar ações.

O sistema pode passar para o estado Fora do mercado (final) seguindo as mesmas condições relacionadas com o estado comprado.

Fora do mercado (fim): Este é o único estado em que não existe uma atuação do Comitê de Decisão. O sistema passa para esse estado em três situações:

- Quando a janela de avaliação termina.
- Se atingiu o *stop loss*.
- Se atingiu o *stop gain*.

O sistema permanecerá nesse estado até o final da janela de avaliação.

Algorithm 10 Pseudocódigo do Sistema de Negociação

```

function TRADINGMODULE(PopEnsemble, TradingWindow, VarTrading)
  InitialDate  $\leftarrow$  getDataInitial(TradingWindow)
  FinalDate  $\leftarrow$  getDataFinal(TradingWindow)
  for CurrentDay = InitialDate to FinalDate do
    [ExpBuy, ExpSell]  $\leftarrow$  Ensemble(PopEnsemble, CurrentDay)
    BuySell  $\leftarrow$  xor(ExpBuy, ExpSell)
    switch VarTrading.Status do
      case 0 ▷ 0-out

        if BuySell then

          if ExpBuy then
            VarTrading  $\leftarrow$  Buying(VarTrading, CurrentDay)
          else
            VarTrading  $\leftarrow$  Selling(VarTrading, CurrentDay)
          end if
        end if
      case 1 ▷ 1-purchased

        if and(BuySell, ExpSell) then
          VarTrading  $\leftarrow$  Selling(VarTrading, CurrentDay)
        end if
      case 2 ▷ 2-sold

        if and(BuySell, ExpBuy) then
          VarTrading  $\leftarrow$  Buying(VarTrading, CurrentDay)
        end if
      R  $\leftarrow$  Account(VarTrading, CurrentDay)
      if (R > StopGain) or (R < StopLoss) then
        Break
      end if
    end for
    VarTrading  $\leftarrow$  Closure(VarTrading, CurrentDay)
    Return (VarTrading)
  end function

```

Onde:

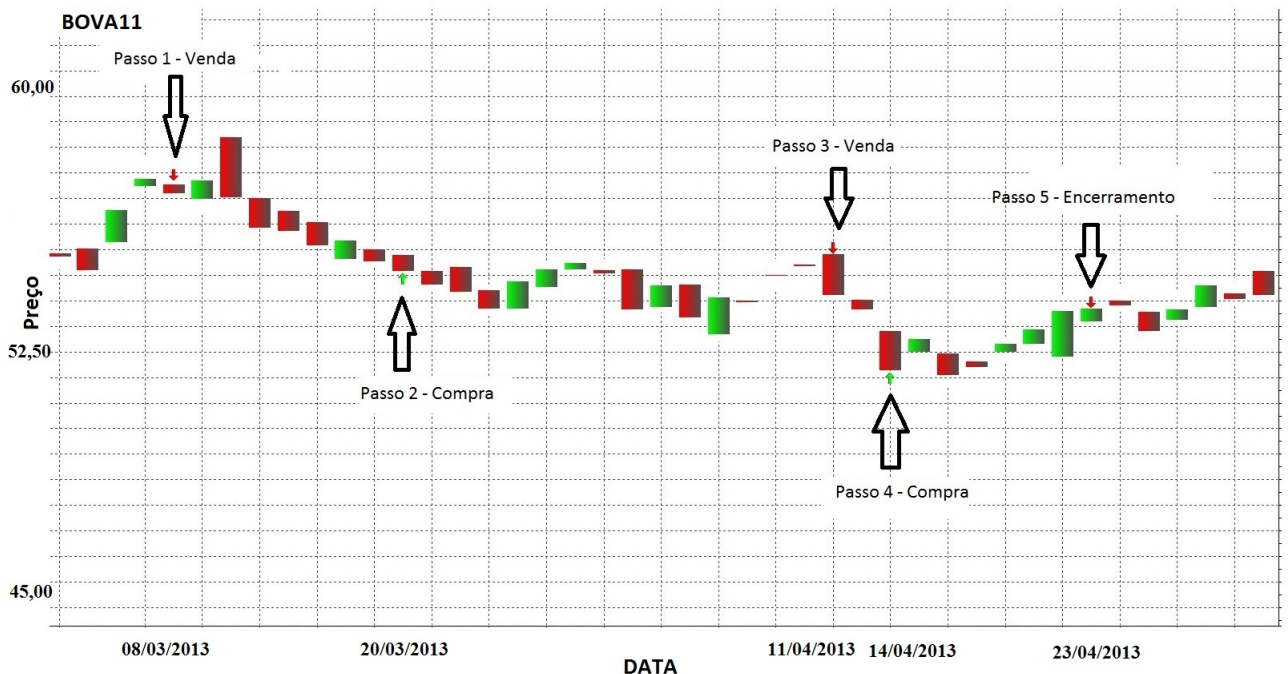
- **getDataInitial**: esta função retorna à data inicial de uma série temporal;
- **getDataFinal**: esta função retorna à data final de uma série temporal;
- **Ensemble**: este algoritmo está detalhado no pseudocódigo 9 e implementa o Comitê de Decisão; cada conjunto de regras é avaliado dia a dia, e a decisão de compra ou de venda é feita utilizando a votação por maioria simples;

- **Buying:** este procedimento efetua a compra de ações e atualiza os status da variável *VarTrading*;
- **Selling:** este algoritmo efetua a venda de ações e atualiza os status da variável *VarTrading*;
- **Account:** este procedimento contabiliza os ganhos e perdas dia a dia; ele é necessário para encerrar as operações em caso de perda máxima estipulada pelo usuário (*Stop Loss*) ou ganho máximo (*Stop Gain*);
- **Closure:** este algoritmo encerra as operações ao final de uma janela de tempo e atualiza a variável *VarTrading*.

4.2.7 Exemplo de funcionamento do módulo de negociação do SisSM

Nesta subseção, é demonstrado o funcionamento do SisSM com foco principal no módulo de negociação (compra e a venda de ações). Para tanto, foi escolhido o papel *BOVA11* em uma janela de 32 dias úteis de negociação. Na figura 26 e na tabela 6, são apresentados um exemplo de funcionamento desse módulo e detalhadas as operações de compra, venda e fechamento.

Figura 26 – BOVA11, exemplo do sistema de negociação



Fonte: Elaborada pelo autor

Regras de compra e venda:

1. A alavancagem pode ser feita em até uma vez o valor do saldo disponível. No passo 1 da tabela 6, o capital inicial é R\$ 1.000,00, mas pode-se fazer uma venda, alugando as ações

Tabela 6 – Exemplo do Sistema de Negociação com o BOVA11

Passo	Preço de Mercado	Posição Inicial	Posição Final	Evolução dos Ativos	Dinheiro em Caixa
1	R\$57,15	Parado	Vendido	R\$1.000,00	R\$2.000,00
2	R\$54,86	Vendido	Comprado	R\$1.040,50	R\$0,00
3	R\$54,15	Comprado	Vendido	R\$1.026,60	R\$2.053,00
4	R\$51,95	Vendido	Comprado	R\$1.067,57	R\$0,00
5	R\$53,69	Comprado	Parado	R\$1.103,32	R\$1.103,32

Fonte: Elaborada pelo autor

no valor de R\$ 1.000,00. Assim:

Ativos = Capital + Total de Ações

Total de Ações Vendidas = Ativos/Valor das ações

Total de Ações Vendidas = R\$ 1.000,00/57,15

Total de Ações Vendidas = 17,49

2. Toda compra é realizada com o dinheiro disponível. No passo 2 da tabela 6, o dinheiro disponível era R\$ 2.000,00. Assim:

Total de Ações Compradas = Dinheiro em Caixa/Preço de Mercado

Total de Ações Compradas = R\$2.000,00/54,86

Total de Ações Compradas = 36,45

Total de Ações = Total de Ações Compradas - Total de Ações Vendidas

Total de Ações = 36,45 - 17,49

Total de Ações = 18,96

A seguir é apresentado o fluxo de execução detalhado desse exemplo, segundo os passos mostrados na tabela 6 e na figura 26.

- Passo 1 - No dia 08/03/2013, o capital inicial era de R\$ 1.000,00. O sistema indicou venda, e o preço da ação era R\$ 57,15. Foram feitos o aluguel e a venda da ação. Contabilizou-se R\$ 2.000,00 em caixa, ainda devendo 17,49 ações.
- Passo 2 - No dia 20/03/2013, o sistema indicou uma compra com o preço da ação em R\$ 54,86. Foi feita a compra de 36,54 ações e contabilizou-se um saldo positivo de 18,96 ações, com R\$ 0,00 em caixa. Os ativos evoluíram para R\$ 1.040,50.
- Passo 3 - No dia 11/04/2013, o sistema indicou uma venda com o preço da ação em R\$ 54,15. Foi feita a venda de 37,92 ações e contabilizou-se um saldo negativo de -18,96 ações com R\$ 2.053,00 em caixa. Os ativos evoluíram para R\$ 1.026,60.

- Passo 4 - No dia 14/04/2013, o sistema indicou uma compra com o preço da ação em R\$ 51,95. Foi feita a compra de 39,51 ações e contabilizou-se um saldo positivo de 20,55 ações, com R\$ 0,00 em caixa. Os ativos evoluíram para R\$ 1.067,57.
- Passo 5 – Por fim, no dia 23/04/2013, é feito o encerramento da operação. Foi feita a venda das ações e apurado o capital. Nesse caso, obteve-se uma evolução de R\$ 1.000,00 para R\$1.103,32.

É importante destacar que não foram tratados os custos operacionais e nem o fracionamento em lote das ações.

4.2.8 Exemplo de funcionamento do SisSM

Esta subseção tem o objetivo de demonstrar o funcionamento do SisSM com foco principal na construção do Comitê de Decisão. Na sequência são apresentadas 16 tabelas, sendo que dessas 16, duas (7 e 8) são referentes aos parâmetros utilizados e às regras selecionadas na “seleção de regras fase I e II” do SisSM; e as 14 tabelas restantes: 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 e 22 detalham passo a passo o funcionamento do sistema.

Optou-se por apresentar uma pequena parcela dos testes realizados. Sendo assim, foi escolhida a ação da companhia CEMIG-MG, entre o período de 05/02/2013 a 28/05/2013, cujos parâmetros estão descritos na tabela 7.

Tabela 7 – Parâmetros adotados no exemplo de funcionamento do SisSM

Parâmetros	Valores
Ação	CMIG4
t_0	05.02.2013
t_{final}	28.05.2013
Δ_F (dias)	30
Δ_T (dias)	30
Δ_E (dias)	10
Capital Inicial	R\$ 1.000,00
N_a	50
N_b	5

Fonte: Elaborada pelo autor

Na tabela 8, é apresentada uma codificação das regras utilizadas, sendo que RC_i representa a (i) ésima regra de compra e RV_i representa a (i) ésima regra de venda. Esse formato foi necessário para auxiliar na visualização das iterações do sistema que são apresentadas nas tabelas seguintes. É importante ressaltar que essas regras apresentadas na tabela 8 são um subconjunto das regras mostradas nas tabelas 4 e 5. Esse subconjunto foi selecionado nas etapas de seleção de regras fase I e seleção de regras fase II, conforme mostrado na subseção 4.2.1.

Tabela 8 – Regras lógicas baseadas em indicadores numéricos da AT selecionadas na fase II do SisSM

IndC	Regra de Compra
RC08	$(WR_{21}(i) < -90)$
RC09	$(E_{14} < 20)$
RC12	$(IFR_2 < 10)$
RC18	$and(OBV(i) > OBV(i-1), PrecoAtivo(i) > PrecoAtivo(i-1))$
RC27	$(PrecoAtivo(i) < MME_{21})$
RC28	$(PrecoAtivo(i-1) < MME_{21})$
RC29	$(PrecoAtivo(i-2) < MME_{21})$
RC31	$(PrecoAtivo(i) < MME_5)$
RC32	$(PrecoAtivo(i-1) < MME_6)$
RC33	$(PrecoAtivo(i-1) < MME_7)$
RC34	$(PrecoAtivo(i-2) < MME_8)$
IndV	Regra de Venda
RV07	$(WR_{14}(i) > -10)$
RV08	$(WR_{21}(i) > -10)$
RV09	$(E_{14}(i) > 90)$
RV12	$(IRF_2(i) > 97)$
RV26	$(PROC_{12}(i) > 10)$

Fonte: Elaborada pelo autor

As tabelas 9, 11, 13, 15, 17, 19 e 21 têm o objetivo principal de apresentar o retorno financeiro (evolução do capital) do SisSM, período a período. Segue a descrição de cada um dos campos dessas tabelas:

- P_i : é a janela de investimento do (i) ésimio período e neste exemplo foram considerados 7 períodos de 10 dias úteis cada (Δ_E);
- RetFin: consolida o retorno total nos períodos testados;
- Evolução do Capital: este campo mostra a evolução do capital utilizado pelo SisSM, período a período e o seu valor inicial é de R\$ 1.000,00;
- Retorno (Período): este campo mostra a evolução do capital em porcentagem;
- Valorização da Ação: este campo mostra a valorização da ação em porcentagem período a período;
- Comitê de Decisão: este campo é fundamental para a compreensão da formação do Comitê de Decisão e nele são apresentadas as soluções que compõem o *emsemble*.

As tabelas 10, 12, 14, 16, 18, 20 e 22 têm o objetivo principal de apresentar a evolução das funções objetivas F_{a1} 4.1 e F_{a2} 4.2, período a período. Segue a descrição de cada um de seus campos:

- F_{a1} : é uma função objetivo descrita na equação 4.1, que tem a finalidade de auxiliar na seleção das regras que irão compor o Comitê de Decisão; é aplicada em cada solução e armazena o retorno financeiro absoluto acumulado ao longo de todos os períodos de negociação;
- F_{a2} : é uma função objetivo descrita na equação 4.2, que tem a finalidade de auxiliar na seleção das regras que irão compor o Comitê de Decisão; é aplicada em cada solução e soma +1, se o retorno no período avaliado for positivo, e subtrai -1, se o retorno no período avaliado for negativo;
- RC: regra de compra, indexada à tabela 8;
- RV: regra de venda, indexada à tabela 8;
- P_i : é a janela de investimento do (i) énesimo período e, neste exemplo, foram considerados 7 períodos de 10 dias úteis cada (Δ_E);
- Ret: consolida o retorno total nos períodos testados.

Nesta subseção, o SisSM é apresentado período a período. Para tanto, cada par de tabelas apresentado a seguir representa uma iteração do SisSM, cujo período possui o tamanho Δ_E e uma tabela mostrando a evolução do capital; finalmente, outra tabela mostrando o desempenho das soluções por meio das funções objetivo F_{a1} e F_{a2} .

A partir da segunda iteração, a seleção das soluções é tratada como um problema de maximização biobjetivo. Sendo assim, é utilizado o algoritmo *Fast Non Dominated Sorting* (DEB et al., 2002) para escolher as soluções não dominadas.

As linhas das tabelas preenchidas com tom cinza indicam a solução escolhida para compor o comitê de decisão na iteração corrente. Também é importante observar que, a cada iteração, são acrescentadas N_b soluções a serem avaliadas. Na sequência, segue o detalhamento das 7 iterações:

1. Iteração I, tabelas (9 e 10): na primeira iteração, conforme o algoritmo 6 apresentado na subseção 4.2.2, são selecionadas as N_b soluções mostradas na tabela 10. Portanto, são selecionadas as soluções 1,2,3,4 e 5 para compor o comitê de decisão. O capital inicial é de R\$ 1.000,00, a valorização da ação é de 0,90% no período e o retorno do SisSM é de 6,23%;
2. Iteração II, tabelas (11 e 12): a partir da segunda iteração, conforme o pseudocódigo 6 apresentado na subseção 4.2.2, é utilizado o algoritmo *Fast Non Dominated Sorting* 4. Portanto, a solução 4, apresentada na tabela 12, é selecionada. Essa é uma solução não dominada, tendo a valorização da ação de 11,73% e o retorno do SisSM de 2,12%;

3. Iteração III, tabelas (13 e 14): na terceira iteração, a solução 4 apresentada na tabela 14 é novamente selecionada, dado o seu desempenho. Ela continua sendo uma solução não dominada, tendo a valorização da ação de -11,73% e o retorno do SisSM de 12,41%;
4. Iteração IV, tabelas (15 e 16): na quarta iteração, a solução 4 apresentada na tabela 16 é novamente selecionada, dado o seu desempenho. Ela continua sendo uma solução não dominada, tendo a valorização da ação de 6,47% e o retorno do SisSM de 6,47%;
5. Iteração V, tabelas (17 e 18): na quinta iteração, a solução 4 apresentada na tabela 18 é novamente selecionada, dado o seu desempenho. Ela continua sendo uma solução não dominada, tendo a valorização da ação de 4,71% e o retorno do SisSM de 6,86%;
6. Iteração VI, tabelas (19 e 20): na sexta iteração, a solução 4 apresentada na tabela 20 é novamente selecionada, dado o seu desempenho. Esta continua sendo uma solução não dominada, tendo a valorização da ação de 3,28% e o retorno do SisSM de 0,47%;
7. Iteração VII, tabelas (21 e 22): finalmente, na sétima iteração, as soluções 4, 13 e 14 são as soluções não dominadas selecionadas, tendo a valorização da ação de 7,23% e o retorno do SisSM de 11,96%. Sendo assim, o SisSM encerra com um retorno financeiro, contabilizando todos os períodos de 56,06%, enquanto que a ação se valorizou nesses mesmos períodos e atingiu 23,08%.

Nesses exemplos é demonstrado que, em determinadas situações, o sistema proposto pode conseguir lucro, não só com o ativo em valorização, mas também no sentido de desvalorização, operando “na ponta comprada e na ponta vendida” do mercado. Especificamente na terceira iteração (ver a tabela 13), é possível observar que a ação desvalorizou -11,68%, enquanto que o SisSM obteve um retorno financeiro de 12,41%. Nesse exemplo, o sistema proposto conseguiu o retorno por meio de uma operação de venda a descoberto.

Essa operação é detalhada na figura 27 e na tabela 23. Na figura 27, está ilustrada a operação de venda a descoberto, em que dois indicadores da AT utilizados pela solução 4 estão representados, sendo: a MME e o WR. Nesse exemplo, foi acionada a regra de venda (RV07) ($WR_{14}(i) > -10$) no dia 18 de março de 2013 e encerrada a operação no dia 25 de março de 2013. A ação foi vendida por R\$ 13.14, no dia 18 de março de 2013, e comprada por R\$ 11.38, no dia 25 de março, obtendo-se uma rentabilidade de 12,41% e encerrando-se, assim, a janela de negociação da 3ª iteração.

Os campos da tabela 23 são descritos a seguir:

- **Data:** data em que a operação foi realizada;
- **Ação:** preço da ação;
- **P_Inicial:** posição inicial do SisSM antes de realizar uma operação;

- **Operação:** comando executado pelo SisSM (compra ou venda);
- **P_Final:** posição final do SisSM depois de realizar uma operação.

Tabela 9 – Exemplo de funcionamento do SisSM, evolução do capital, iteração I

	<i>PI</i>	<i>RetFinal</i>
Evolução do Capital (R\$)	R\$ 1.000,00	R\$ 1.062,30
Retorno (Período)(%)	6,23	6,23
Valorização da Ação(%)	0,90	0,90
Comitê de Decisão	1*2*3*4*5	****

Fonte: Elaborada pelo autor

Tabela 10 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração I

<i>Ind</i>	<i>F_{a1}(%)</i>	<i>F_{a2}</i>	<i>RC</i>	<i>RV</i>	<i>PI(%)</i>	<i>Ret(%)</i>
1	***	***	RC08	RV07	2,57	2,57
2	***	***	RC08	RV08	0,00	0,00
3	***	***	RC08	RV09	2,57	2,57
4	***	***	RC28	RV07	6,23	6,23
5	***	***	RC28	RV08	0,90	0,90

Fonte: Elaborada pelo autor

Tabela 11 – Exemplo de funcionamento do SisSM, evolução do capital, iteração II

	<i>P1</i>	<i>P2</i>	<i>RetFinal</i>
Evolução do Capital (R\$)	R\$ 1.000,00	R\$ 1.062,30	R\$ 1.084,79
Retorno (Período)(%)	6,23	2,12	8,48
Valorização da Ação(%)	0,90	11,73	14,40
Comitê de Decisão	1*2*3*4*5	4*	****

Fonte: Elaborada pelo autor

Tabela 12 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração II

<i>Ind</i>	<i>F_{a1}(%)</i>	<i>F_{a2}</i>	<i>RC</i>	<i>RV</i>	<i>PI(%)</i>	<i>P2(%)</i>	<i>Ret(%)</i>
1	2,57	1	RC08	RV07	2,57	-4,49	-1,92
2	0,00	0	RC08	RV08	0,00	-4,49	-4,49
3	2,57	1	RC08	RV09	2,57	-4,49	-1,92
4	6,23	1	RC28	RV07	6,23	2,12	8,35
5	0,90	1	RC28	RV08	0,90	2,12	3,02
6			RC31	RV09		-0,17	-0,17
7			RC28	RV07		-0,17	-0,17
8			RC28	RV07		2,12	2,12
9			RC28	RV09		2,12	2,12
10			RC27	RV07		-4,49	-4,49

Fonte: Elaborada pelo autor

Tabela 13 – Exemplo de funcionamento do SisSM, evolução do capital, iteração III

	<i>PI</i>	<i>P2</i>	<i>P3</i>	<i>RetFinal</i>
Evolução do Capital (R\$)	R\$ 1.000,00	R\$ 1.062,30	R\$ 1.084,79	R\$ 1.219,38
Retorno (Período)(%)	6,23	2,12	12,41	21,94
Valorização da Ação(%)	0,9	11,73	-11,68	1,08
Comitê de Decisão	1*2*3*4*5*	4*	4*	*****

Fonte: Elaborada pelo autor

Tabela 14 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração III

<i>Ind</i>	<i>F_{a1}(%)</i>	<i>F_{a2}</i>	<i>RC</i>	<i>RV</i>	<i>PI(%)</i>	<i>P2(%)</i>	<i>P3(%)</i>	<i>Ret(%)</i>
1	-1,92	0	RC08	RV07	2,57	-4,49	15,10	13,17
2	-4,49	-1	RC08	RV08	0,00	-4,49	15,10	10,60
3	-1,92	0	RC08	RV09	2,57	-4,49	15,10	13,17
4	8,35	2	RC28	RV07	6,23	2,12	12,41	20,75
5	3,02	2	RC28	RV08	0,90	2,12	12,41	15,43
6	-0,17	-1	RC31	RV07	0,00	-0,17	10,22	10,05
7	-0,17	-1	RC31	RV09	0,00	-0,17	10,22	10,05
8	2,12	1	RC28	RV07	0,00	2,12	12,41	14,52
9	2,12	1	RC28	RV09	0,00	2,12	12,41	14,52
10	-4,49	-1	RC27	RV07	0,00	-4,49	15,10	10,60
11			RC27	RV07	0,00	0,00	15,10	15,10
12			RC27	RV09	0,00	0,00	15,10	15,10
13			RC31	RV07	0,00	0,00	20,46	20,46
14			RC31	RV09	0,00	0,00	20,46	20,46
15			RC27	RV07	0,00	0,00	12,41	12,41

Fonte: Elaborada pelo autor

Tabela 15 – Exemplo de funcionamento do SisSM, evolução do capital, iteração IV

	<i>P1</i>	<i>P2</i>	<i>P3</i>	<i>P4</i>	<i>RetFinal</i>
Evolução do Capital (R\$)	R\$ 1.000,00	R\$ 1.062,30	R\$ 1.084,79	R\$ 1.219,38	R\$ 1.298,32
Retorno (Período)(%)	6,23	2,12	12,41	6,47	29,83
Valorização da Ação(%)	0,90	11,73	-11,68	6,47	9,89
Comitê de Decisão	1*2*3*4*5*	4*	4*	4*	****

Fonte: Elaborada pelo autor

Tabela 16 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração IV

<i>Ind</i>	$F_{a1}(\%)$	F_{a2}	<i>RC</i>	<i>RV</i>	<i>P1</i> (%)	<i>P2</i> (%)	<i>P3</i> (%)	<i>P4</i> (%)	<i>Ret</i> (%)
1	13,17	1	RC08	RV07	2,57	-4,49	15,10	0,00	13,17
2	10,60	0	RC08	RV08	0,00	-4,49	15,10	0,00	10,60
3	13,17	1	RC08	RV09	2,57	-4,49	15,10	0,00	13,17
4	20,75	3	RC28	RV07	6,23	2,12	12,41	6,47	27,23
5	15,43	3	RC28	RV08	0,90	2,12	12,41	6,47	21,90
6	10,05	0	RC31	RV07	0,00	-0,17	10,22	6,47	16,52
7	10,05	0	RC31	RV09	0,00	-0,17	10,22	6,47	16,52
8	14,52	2	RC28	RV07	0,00	2,12	12,41	6,47	21,00
9	14,52	2	RC28	RV09	0,00	2,12	12,41	6,47	21,00
10	10,60	0	RC27	RV07	0,00	-4,49	15,10	6,47	17,08
11	15,10	1	RC27	RV07	0,00	0,00	15,10	6,47	21,57
12	15,10	1	RC27	RV09	0,00	0,00	15,10	6,47	21,57
13	20,46	1	RC31	RV07	0,00	0,00	20,46	6,47	26,93
14	20,46	1	RC31	RV09	0,00	0,00	20,46	6,47	26,93
15	12,41	1	RC27	RV07	0,00	0,00	12,41	6,47	18,88
16			RC27	RV07	0,00	0,00	0,00	6,47	6,47
17			RC27	RV09	0,00	0,00	0,00	6,47	6,47
18			RC09	RV08	0,00	0,00	0,00	6,47	6,47
19			RC27	RV08	0,00	0,00	0,00	6,47	6,47
20			RC27	RV07	0,00	0,00	0,00	6,47	6,47

Fonte: Elaborada pelo autor

Tabela 17 – Exemplo de funcionamento do SisSM, evolução do capital, iteração V

	Evolução do Capital (R\$)	Retorno (%)	Ação(%)	Comitê de Decisão
P1	R\$ 1.000,00	6,23	0,90	1*2*3*4*5*
P2	R\$ 1.062,30	2,12	11,73	4*
P3	R\$ 1.084,79	12,41	-11,68	4*
P4	R\$ 1.219,38	6,47	6,47	4*
P5	R\$ 1.298,32	6,86	4,71	4*
RetFinal	R\$ 1.387,35	12,37	0,90	*****

Fonte: Elaborada pelo autor

Tabela 18 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração V

Ind	$F_{a1}(\%)$	F_{a2}	RC	RV	P1(%)	P2(%)	P3(%)	P4(%)	P5(%)	Ret(%)
1	13,17	0	RC08	RV07	2,57	-4,49	15,10	0,00	0,40	13,57
2	10,60	0	RC08	RV08	0,00	-4,49	15,10	0,00	-2,43	8,17
3	13,17	0	RC08	RV09	2,57	-4,49	15,10	0,00	0,40	13,57
4	27,23	4	RC28	RV07	6,23	2,12	12,41	6,47	6,86	34,08
5	21,90	4	RC28	RV08	0,90	2,12	12,41	6,47	0,98	22,88
6	16,52	1	RC31	RV07	0,00	-0,17	10,22	6,47	4,98	21,50
7	16,52	1	RC31	RV09	0,00	-0,17	10,22	6,47	4,98	21,50
8	21,00	3	RC28	RV07	0,00	2,12	12,41	6,47	6,86	27,85
9	21,00	3	RC28	RV09	0,00	2,12	12,41	6,47	6,86	27,85
10	17,08	1	RC27	RV07	0,00	-4,49	15,10	6,47	6,72	23,80
11	21,57	2	RC27	RV07	0,00	0,00	15,10	6,47	6,72	28,29
12	21,57	2	RC27	RV09	0,00	0,00	15,10	6,47	6,72	28,29
13	26,93	2	RC31	RV07	0,00	0,00	20,46	6,47	6,72	33,65
14	26,93	2	RC31	RV09	0,00	0,00	20,46	6,47	6,72	33,65
15	18,88	2	RC27	RV07	0,00	0,00	12,41	6,47	6,86	25,74
16	6,47	1	RC27	RV07	0,00	0,00	0,00	6,47	6,72	13,19
17	6,47	1	RC27	RV09	0,00	0,00	0,00	6,47	6,72	13,19
18	6,47	1	RC09	RV08	0,00	0,00	0,00	6,47	-2,43	4,05
19	6,47	1	RC27	RV08	0,00	0,00	0,00	6,47	0,85	7,32
20	6,47	1	RC27	RV07	0,00	0,00	0,00	6,47	6,86	13,33
21			RC33	RV07	0,00	0,00	0,00	0,00	6,86	6,86
22			RC33	RV09	0,00	0,00	0,00	0,00	6,86	6,86
23			RC33	RV08	0,00	0,00	0,00	0,00	6,86	6,86
24			RC18	RV12	0,00	0,00	0,00	0,00	4,98	4,98
25			RC34	RV07	0,00	0,00	0,00	0,00	4,98	4,98

Fonte: Elaborada pelo autor

Tabela 19 – Exemplo de funcionamento do SisSM, evolução do capital, iteração VI

	Evolução do Capital (R\$)	Retorno (%)	Ação(%)	Comitê de Decisão
P1	R\$ 1.000,00	6,23	0,90	1*2*3*4*5*
P2	R\$ 1.062,30	2,12	11,73	4*
P3	R\$ 1.084,79	12,41	-11,68	4*
P4	R\$ 1.219,38	6,47	6,47	4*
P5	R\$ 1.298,32	6,86	4,71	4*
P6	R\$ 1.387,35	0,47	3,28	4*
RetFinal	R\$ 1.393,88	39,39	13,99	****

Fonte: Elaborada pelo autor

Tabela 20 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração VI

Ind	$F_{a1}(\%)$	F_{a2}	RC	RV	PI(%)	P2(%)	P3(%)	P4(%)	P5(%)	P6(%)	Ret(%)
1	13,57	1	RC08	RV07	2,57	-4,49	15,10	0,00	0,40	0,47	14,04
2	8,17	-1	RC08	RV08	0,00	-4,49	15,10	0,00	-2,43	0,47	8,64
3	13,57	1	RC08	RV09	2,57	-4,49	15,10	0,00	0,40	0,47	14,04
4	34,08	5	RC28	RV07	6,23	2,12	12,41	6,47	6,86	0,47	34,56
5	22,88	5	RC28	RV08	0,90	2,12	12,41	6,47	0,98	0,47	23,35
6	21,50	2	RC31	RV07	0,00	-0,17	10,22	6,47	4,98	0,47	21,97
7	21,50	2	RC31	RV09	0,00	-0,17	10,22	6,47	4,98	0,47	21,97
8	27,85	4	RC28	RV07	0,00	2,12	12,41	6,47	6,86	0,47	28,33
9	27,85	4	RC28	RV09	0,00	2,12	12,41	6,47	6,86	0,47	28,33
10	23,80	3	RC27	RV07	0,00	-4,49	15,10	6,47	6,72	0,47	24,27
11	28,29	3	RC27	RV07	0,00	0,00	15,10	6,47	6,72	0,47	28,76
12	28,29	3	RC27	RV09	0,00	0,00	15,10	6,47	6,72	0,47	28,76
13	33,65	3	RC31	RV07	0,00	0,00	20,46	6,47	6,72	4,26	37,91
14	33,65	3	RC31	RV09	0,00	0,00	20,46	6,47	6,72	4,26	37,91
15	25,74	3	RC27	RV07	0,00	0,00	12,41	6,47	6,86	0,47	26,21
16	13,19	2	RC27	RV07	0,00	0,00	0,00	6,47	6,72	0,47	13,67
17	13,19	2	RC27	RV09	0,00	0,00	0,00	6,47	6,72	0,47	13,67
18	4,05	0	RC09	RV08	0,00	0,00	0,00	6,47	-2,43	0,47	4,52
19	7,32	2	RC27	RV08	0,00	0,00	0,00	6,47	0,85	0,47	7,80
20	13,33	2	RC27	RV07	0,00	0,00	0,00	6,47	6,86	0,47	13,80
21	6,86	1	RC33	RV07	0,00	0,00	0,00	0,00	6,86	2,41	9,27
22	6,86	1	RC33	RV09	0,00	0,00	0,00	0,00	6,86	2,41	9,27
23	6,86	1	RC33	RV08	0,00	0,00	0,00	0,00	6,86	2,41	9,27
24	4,98	1	RC18	RV12	0,00	0,00	0,00	0,00	4,98	2,41	7,39
25	4,98	1	RC34	RV07	0,00	0,00	0,00	0,00	4,98	0,00	4,97
26			RC33	RV12	0,00	0,00	0,00	0,00	0,00	1,47	1,47
27			RC33	RV07	0,00	0,00	0,00	0,00	0,00	2,41	2,41
28			RC33	RV09	0,00	0,00	0,00	0,00	0,00	2,41	2,41
29			RC32	RV26	0,00	0,00	0,00	0,00	0,00	3,28	3,28
30			RC12	RV26	0,00	0,00	0,00	0,00	0,00	3,28	3,28

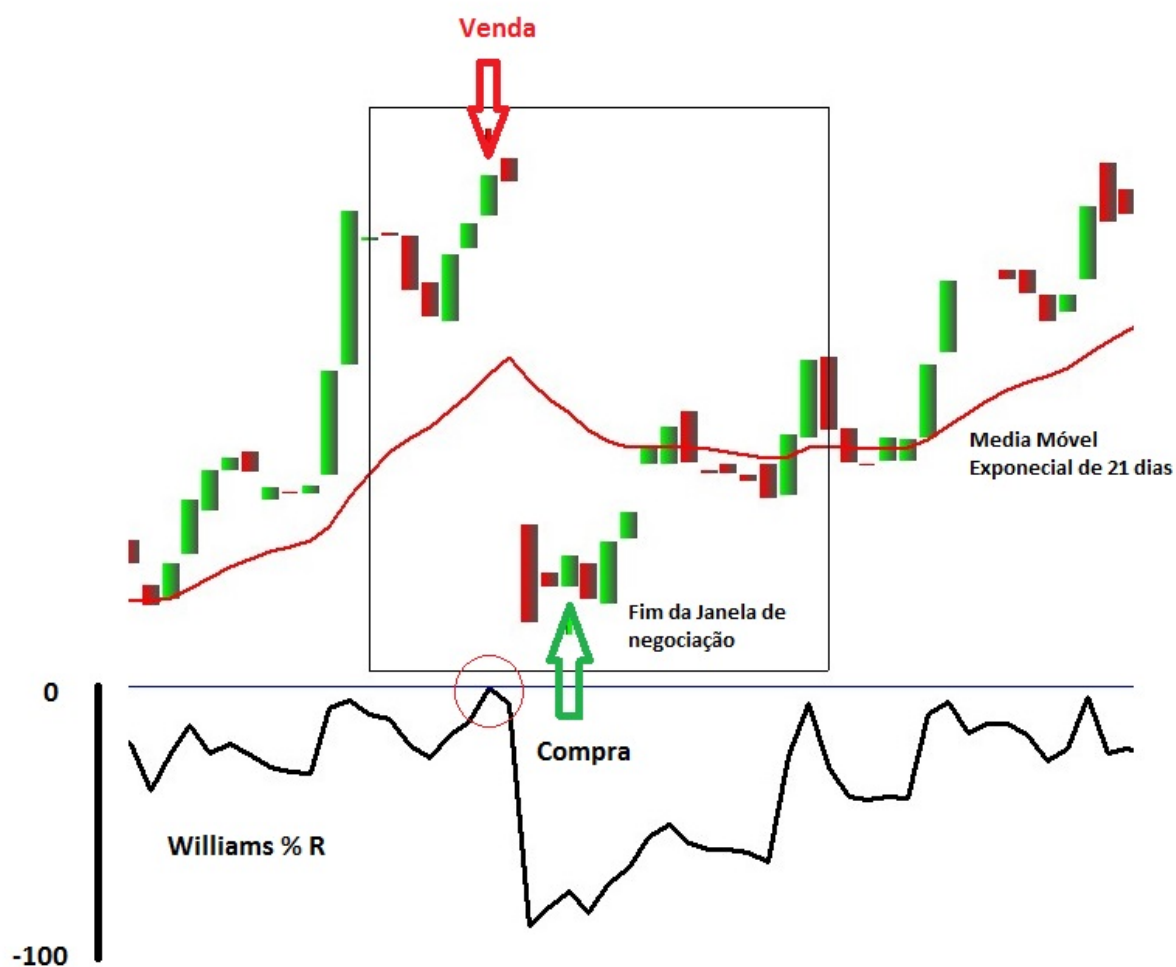
Fonte: Elaborada pelo autor

Tabela 21 – Exemplo de funcionamento do SisSM, evolução do capital, iteração VII

	Evolução do Capital (R\$)	Retorno (%)	Ação(%)	Comitê de Decisão
<i>P1</i>	R\$ 1.000,00	6,23	0,90	1*2*3*4*5*
<i>P2</i>	R\$ 1.062,30	2,12	11,73	4*
<i>P3</i>	R\$ 1.084,79	12,41	-11,68	4*
<i>P4</i>	R\$ 1.219,38	6,47	6,47	4*
<i>P5</i>	R\$ 1.298,32	6,86	4,71	4*
<i>P6</i>	R\$ 1.387,35	0,47	3,28	4*
<i>P7</i>	R\$ 1.393,88	11,96	7,34	4*13*14*
<i>RetFinal</i>	R\$ 1.560,57	56,06	23,08	****

Fonte: Elaborada pelo autor

Figura 27 – Exemplo de venda a descoberto, Ação - CMIG4



Fonte: Elaborada pelo autor

Tabela 22 – Exemplo de funcionamento do SisSM, acompanhamento das funções objetivo F_{a1} 4.1 e F_{a2} 4.2, iteração VII

<i>Ind</i>	<i>F_{a1}</i> (%) (%)	<i>F_{a2}</i>	<i>RC</i>	<i>RV</i>	<i>P1</i> (%)	<i>P2</i> (%)	<i>P3</i> (%)	<i>P4</i> (%)	<i>P5</i> (%)	<i>P6</i> (%)	<i>P7</i> (%)	<i>Ret</i> (%)
1	14,04	2	RC08	RV07	2,57	-4,49	15,10	0,00	0,40	0,47	2,10	16,15
2	8,64	0	RC08	RV08	0,00	-4,49	15,10	0,00	-2,43	0,47	-3,17	5,48
3	14,04	2	RC08	RV09	2,57	-4,49	15,10	0,00	0,40	0,47	2,10	16,15
4	34,56	6	RC28	RV07	6,23	2,12	12,41	6,47	6,86	0,47	2,10	36,66
5	23,35	6	RC28	RV08	0,90	2,12	12,41	6,47	0,98	0,47	-3,17	20,18
6	21,97	3	RC31	RV07	0,00	-0,17	10,22	6,47	4,98	0,47	2,10	24,07
7	21,97	3	RC31	RV09	0,00	-0,17	10,22	6,47	4,98	0,47	2,10	24,07
8	28,33	5	RC28	RV07	0,00	2,12	12,41	6,47	6,86	0,47	2,10	30,43
9	28,33	5	RC28	RV09	0,00	2,12	12,41	6,47	6,86	0,47	2,10	30,43
10	24,27	3	RC27	RV07	0,00	-4,49	15,10	6,47	6,72	0,47	2,10	26,37
11	28,76	4	RC27	RV07	0,00	0,00	15,10	6,47	6,72	0,47	2,10	30,87
12	28,76	4	RC27	RV09	0,00	0,00	15,10	6,47	6,72	0,47	2,10	30,87
13	37,91	4	RC31	RV07	0,00	0,00	20,46	6,47	6,72	4,26	11,96	49,87
14	37,91	4	RC31	RV09	0,00	0,00	20,46	6,47	6,72	4,26	11,96	49,87
15	26,21	4	RC27	RV12	0,00	0,00	12,41	6,47	6,86	0,47	2,10	28,31
16	13,67	3	RC27	RV07	0,00	0,00	0,00	6,47	6,72	0,47	2,10	15,77
17	13,67	3	RC27	RV09	0,00	0,00	0,00	6,47	6,72	0,47	2,10	15,77
18	4,52	1	RC09	RV08	0,00	0,00	0,00	6,47	-2,43	0,47	-3,17	1,35
19	7,80	3	RC27	RV08	0,00	0,00	0,00	6,47	0,85	0,47	-3,17	4,63
20	13,80	3	RC27	RV08	0,00	0,00	0,00	6,47	6,86	0,47	2,10	15,91
21	9,27	2	RC33	RV07	0,00	0,00	0,00	0,00	6,86	2,41	11,96	21,23
22	9,27	2	RC33	RV07	0,00	0,00	0,00	0,00	6,86	2,41	11,96	21,23
23	9,27	2	RC33	RV09	0,00	0,00	0,00	0,00	6,86	2,41	11,96	21,23
24	7,39	2	RC33	RV08	0,00	0,00	0,00	0,00	4,98	2,41	6,94	14,33
25	4,97	0	RC18	RV12	0,00	0,00	0,00	0,00	4,98	0,00	11,54	16,51
26	1,47	2	RC34	RV07	0,00	0,00	0,00	0,00	0,00	1,47	4,71	6,18
27	2,41	2	RC33	RV07	0,00	0,00	0,00	0,00	0,00	2,41	11,96	14,37
28	2,41	2	RC33	RV09	0,00	0,00	0,00	0,00	0,00	2,41	11,96	14,37
29	3,28	2	RC32	RV26	0,00	0,00	0,00	0,00	0,00	3,28	7,34	10,63
30	3,28	2	RC12	RV26	0,00	0,00	0,00	0,00	0,00	3,28	7,34	10,63
31			RC32	RV07	0,00	0,00	0,00	0,00	0,00	0,00	11,96	11,96
32			RC32	RV09	0,00	0,00	0,00	0,00	0,00	0,00	11,96	11,96
33			RC33	RV12	0,00	0,00	0,00	0,00	0,00	0,00	4,71	4,71
34			RC12	RV07	0,00	0,00	0,00	0,00	0,00	0,00	11,96	11,96
35			RC33	RV07	0,00	0,00	0,00	0,00	0,00	0,00	11,96	11,96

Fonte: Elaborada pelo autor

Tabela 23 – Detalhamento da terceira iteração, exemplo de venda a descoberto, Ação - CMIG4

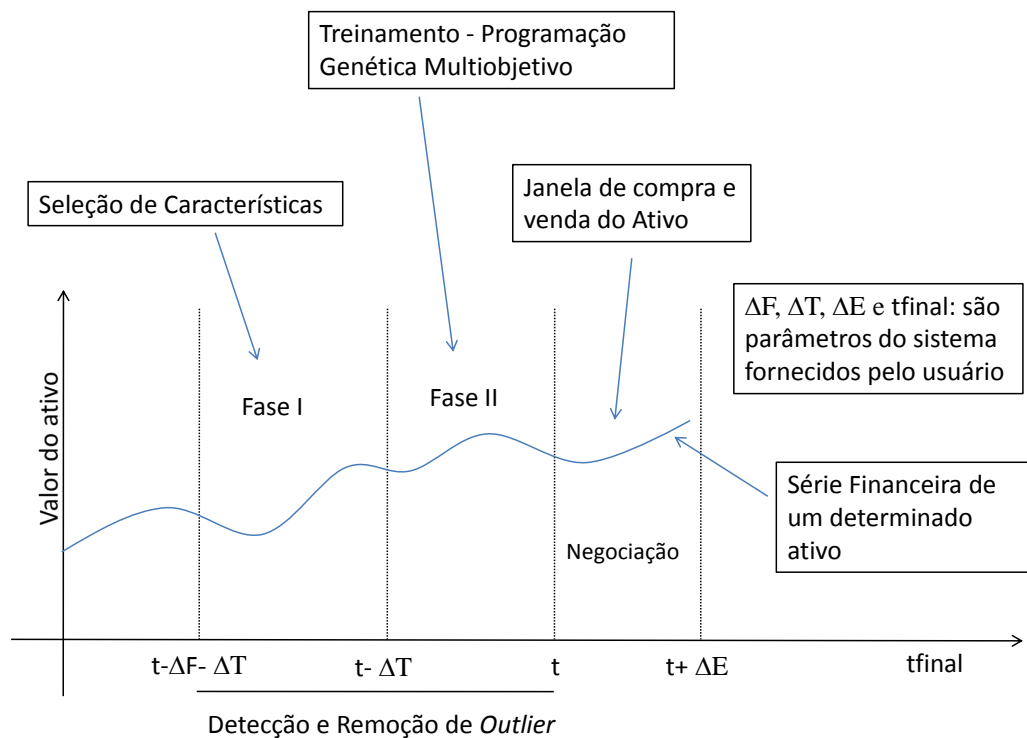
Data	Ação	P_Inicial	Operação	P_Final
18/mar/13	R\$ 13.14	Fora	Venda	Vendido
21/mar/13	R\$ 11.44	Vendido	Compra	Comprado
25/mar/13	R\$ 11.38	Comprado	Venda	Fora

Fonte: Elaborada pelo autor

4.3 Sistema Automatizado de Investimento via Programação Genética Multiobjetivo

O SisPG também trabalha em três janelas móveis de tempo. Na primeira janela, é realizada a seleção de características. Na segunda janela, é realizada a geração de regras via programação genética multiobjetivo e a montagem do Comitê de Decisão. Finalmente, na terceira janela, é feita a negociação das ações. Na figura 28, são esboçadas as fases do sistema.

Figura 28 – Fases do SisPG



Fonte: Elaborada pelo autor

Nas subseções seguintes são detalhadas as cinco etapas do sistema: detecção e remoção de *Outliers*, Seleção de Características, treinamento utilizando uma PG multiobjetivo, o Comitê de Decisão e, por fim, a etapa de negociação.

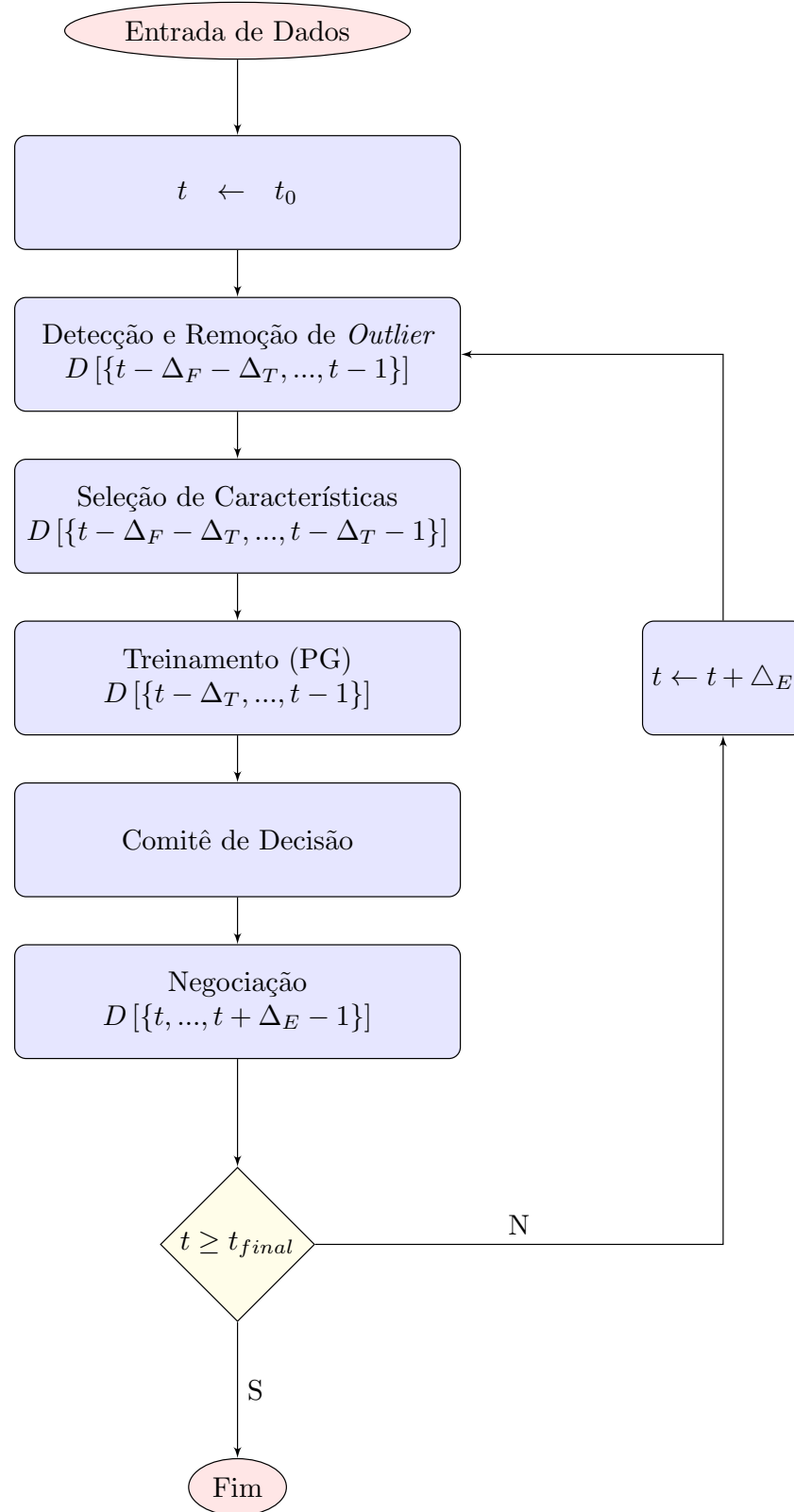
4.3.1 Visão Geral do Sistema do SisPG

O esquema geral do SisPG é apresentado na figura 29. É possível notar que o sistema trabalha com janelas móveis dentro de um laço de repetição, onde a condição de parada é quando a data final (t_{final}) é atingida. Três janelas de tempo são usadas em cada iteração: a janela de Seleção de Características $\{t - \Delta_F - \Delta_T, \dots, t - \Delta_T - 1\}$ (dados passados), a janela de treinamento $\{t - \Delta_T, \dots, t - 1\}$ (dados passados) e a janela de investimentos $\{t, \dots, t + \Delta_E - 1\}$ (dados futuros). O sistema proposto pode ser resumido em cinco operações principais que são realizadas em cada iteração:

1. Remoção de *Outliers* em $D[\{t - \Delta_F - \Delta_T, \dots, t - 1\}]$: *Outliers* são identificados e removidos.
2. Seleção de Características $D[\{t - \Delta_F - \Delta_T, \dots, t - \Delta_T - 1\}]$: os atributos candidatos são avaliados na janela de Seleção de Características. Os N_a melhores atributos são considerados na fase de treinamento, enquanto os restantes são ignorados.
3. Treinamento (PG) $D[\{t - \Delta_T, \dots, t - 1\}]$: nesta etapa, a PG multi-objetivo é aplicada ao intervalo de treinamento para construir regras de compra e regras de venda.
4. Comitê de Decisão: as regras obtidas pela PG são usadas para construir o Comitê de Decisão, que tomará as decisões na etapa de negociação.
5. Negociação $D[\{t, \dots, t + \Delta_E - 1\}]$: finalmente, o Comitê de Decisão é usado para identificar se uma das três operações serão realizadas no final de cada dia na janela de avaliação: (i) comprar; (ii) vender, ou; (iii) manter.

Nas próximas subseções são apresentados os algoritmos do SisPG juntamente com cada uma das etapas descritas. $N_a, t_0, t_{final}, \Delta_T, \Delta_F, \Delta_E$ são parâmetros definidos pelo usuário.

Figura 29 – Esquema Geral do SisPG. Neste esquema: $D[\{t_I, \dots, t_F\}]$ é a série financeira inicial t_I e final t_F ; $\Delta_F > 0$, $\Delta_T > 0$, e $\Delta_E > 0$ tamanho das janelas de Seleção de Características, Treinamento e Negociação, respectivamente



Fonte: Elaborada pelo autor

4.3.2 Algoritmo geral do SisPG

O algoritmo do SisPG é apresentado no pseudocódigo 11.

Algorithm 11 Pseudocódigo do SisPG

```
 $t \leftarrow t_0;$ 
 $PopEnsemble \leftarrow \emptyset;$ 
 $SetAllRules \leftarrow \text{SelectAllRules};$ 
 $VarTrading.Status \leftarrow 0;$ 
 $VarTrading.ProfitLoss \leftarrow 0;$ 
 $VarTrading.Cash \leftarrow InitialCapital;$ 
 $VarTrading.Share \leftarrow 0;$ 
while  $t < t_{final}$  do
     $SlidingWindow \leftarrow \text{SelectData}(FinancialSeries, t - \Delta_F - \Delta_T, t - 1);$ 
     $SlidingWindow \leftarrow \text{RemoveOutlier}(SlidingWindow, F_{low});$ 
     $SlidingWindowI \leftarrow \text{SelectData}(SlidingWindow, t - \Delta_F - \Delta_T, t - \Delta_T - 1);$ 
     $SlidingWindowII \leftarrow \text{SelectData}(SlidingWindow, t - \Delta_T, t - 1);$ 
     $TradingWindow \leftarrow \text{SelectData}(FinancialSeries, t, t + \Delta_E - 1);$ 
     $Rules \leftarrow \text{FeatureSelection}(SlidingWindowI, SetAllRules, N_a);$ 
     $PopEnsemble \leftarrow \text{Training}(SlidingWindowII, Rules);$ 
     $VarTrading \leftarrow \text{TradingModule}(PopEnsemble, TradingWindow, VarTrading);$ 
     $t \leftarrow t + \Delta_E;$ 
end while
 $ProfitLoss \leftarrow VarTrading.ProfitLoss;$ 
```

Onde:

- $t_0, t_{final}, \Delta_F, \Delta_T, \Delta_E, N_a, InitialCapital$: são parâmetros fornecidos pelo usuário;
- **VarTrading**: é uma variável estruturada com quatro campos distintos, de fundamental importância para o controle do SisPG. O campo **VarTrading.Status** pode assumir três valores, que indicam a situação do sistema no *trading*: o valor 0 significa que o sistema está fora do *trading*, o valor 1 significa que o sistema está na situação comprado e o valor 2 significa que o sistema está na situação vendido. O campo **VarTrading.Cash** registra o quanto de valor o sistema tem em caixa para negociar a compra ou a venda de ações. O campo **VarTrading.ProfitLoss** armazena os lucros e prejuízos que o sistema obtém ao longo das N janelas de negociação. O campo **VarTrading.Share** registra a quantidade de ações disponíveis. É importante destacar que esse campo pode ter valor negativo, considerando que é permitido vender ações a descoberto (ELDER, 1993);
- **SelectAllRules**: este procedimento busca todas as regras de compra e de venda apresentadas na tabela;
- **SelectData**: este procedimento seleciona os dados de um período de uma dada série financeira;

- **RemoveOutlier:** dada uma série financeira, este procedimento detecta e remove os *Outlier*; o algoritmo é descrito no pseudocódigo 12 e utiliza como base o método descrito em Cleveland (1981);
- **FeatureSelection:** o objetivo deste algoritmo é selecionar um conjunto (N_a) de melhores atributos para serem utilizados na fase de treinamento. O melhor atributo é uma regra de compra e de venda que obtém o maior retorno financeiro testado em uma dada série financeira; o algoritmo está detalhado no pseudocódigo 13;
- **Training:** este algoritmo é descrito no pseudocódigo 14 e utiliza a programação genética multiobjetivo, combinando os operadores lógicos *AND*, *OR* e *XOR* com indicadores da AT para construir as regras de compra e de venda. É utilizado o algoritmo evolucionário multiobjetivo NSGA-II, conforme proposto por Deb et al. (2002) e descrito em detalhes no capítulo 3, seção 3.5.
- **TradingModule:** este algoritmo está descrito no pseudocódigo 10. Dada uma série financeira, ele realiza as operações de compra, de venda e encerra as operações, quando necessário.

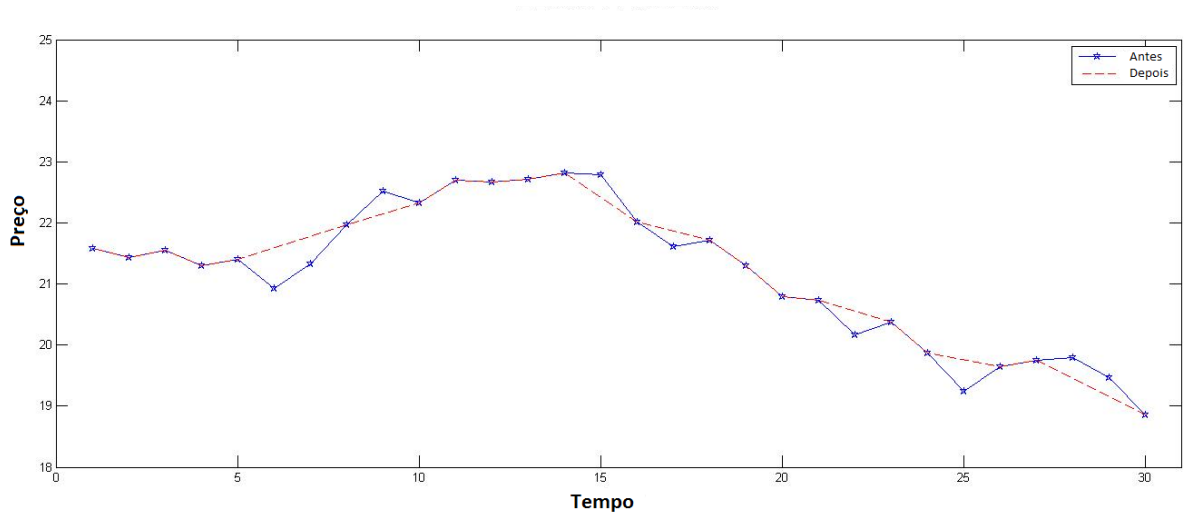
4.3.3 Detecção e Remoção de *Outliers* do SisPG

De acordo com a literatura, “*Outliers* são observações que não seguem o padrão da maioria dos dados”(COOK; HAWKINS, 1990). Em uma série temporal financeira é comum observar variações extremas no valor do ativo. Muitas vezes, essas variações são indicativos de ruídos que podem afetar a generalização dos modelos gerados.

Neste trabalho, é proposto um método automático de Detecção e Remoção de *Outliers*. Este método é baseado no *Locally Weighted Scatterplot Smoothing* (LOWESS) (CLEVELAND, 1981). Segue a descrição do procedimento, cujo pseudocódigo é apresentado no algoritmo 12:

1. Constrói-se uma aproximação suave da curva de preço da ação, usando um método baseado no *Locally Weighted Scatterplot Smoothing* (LOWESS) (CLEVELAND, 1981);
2. Avaliam-se as diferenças entre os dados observados e o ponto respectivo na aproximação suave;
3. Define-se a diferença negativa máxima aceitável como $MaxNeg \leftarrow Q_1 - F_{low}(Q_3 - Q_1)$, no qual Q_1 e Q_3 são o primeiro e o terceiro quartis das diferenças, e F_{low} é um parâmetro definido pelo usuário;
4. Define-se a diferença positiva máxima aceitável como $MaxPos \leftarrow Q_3 + F_{low}(Q_3 - Q_1)$;
5. Remove-se qualquer ponto de treinamento cuja diferença (*dif*) encontra-se numa das seguintes condições:: $dif < MaxNeg$ ou $dif > MaxPos$.

Figura 30 – Exemplo de dados (BBAS3 entre 7 de maio de 2013 e 19 de junho de 2013). Antes (linha contínua preta) e Depois (linha tracejada vermelha) de aplicar o algoritmo de detecção e remoção de *Outlier*



Fonte: Elaborada pelo autor

Basicamente, esse algoritmo considera uma faixa de aceitação com base na distância interquartis. O tamanho dessa faixa é definido pelo parâmetro F_{low} .

Um exemplo da aplicação deste procedimento é apresentado na figura 30, para o preço das ações do Banco do Brasil (BBAS3). É possível notar que ele filtra as variações súbitas do preço das ações.

Algorithm 12 Pseudocódigo do Remove *Outlier*

```

function REMOVEOUTLIER(SlidingWindow,  $F_{low}$ )
     $O_r \leftarrow \text{Cleveland}(\textit{SlidingWindow})$ ;           ▷ output.residual(CLEVELAND, 1981)
     $Q_1 \leftarrow \text{Quartiles\_function}(O_r, 1)$ ;           ▷ primeiro quartil
     $Q_3 \leftarrow \text{Quartiles\_function}(O_r, 3)$ ;           ▷ terceiro quartil
     $minVal \leftarrow Q_1 - F_{low} * (Q_3 - Q_1)$ ;
     $maxVal \leftarrow Q_3 + F_{low} * (Q_3 - Q_1)$ ;
     $List \leftarrow \text{find}((Q_r \geq minVal) \textbf{and} (Q_r \leq maxVal))$ ;
     $NewTrainData \leftarrow \textit{SlidingWindow}(List)$ ;
    return ( $NewTrainData$ )
end function

```

Onde:

- *SlidingWindow*: é a série financeira de onde se pretende remover os *Outlier*;
- F_{low} : parâmetro fornecido pelo usuário;
- **Cleveland**: o algoritmo proposto por Cleveland (1981), que é utilizado para detecção dos *Outliers*.

4.3.4 Seleção de Características do SisPG

No SisPG foram utilizados os mesmos indicadores numéricos da AT utilizados no SisSM, já apresentados na tabela 3. Como já foi explicado anteriormente, estes indicadores foram escolhidos por conterem os quatro grupos principais da AT computadorizada e por serem relativamente fáceis de serem calculados.

Apesar de utilizar os mesmos indicadores técnicos usado no SisSM, no SisPG foi criado um outro conjunto de regras de compra e de venda, apresentados nas tabelas 24 e 25 respectivamente. Essas regras são atributos candidatos para a PG Multiobjetivo e devem ser lidas como o exemplo a seguir:

- Compra :: $MMS_{21}(i) > MMS_{21}(i - 3)$: Se a MMS avaliada para 21 dias, terminando hoje, é maior que a MMS de 21 dias, que terminou há três dias, então compre, sugere uma tendência de alta.
- Venda :: $IFR_{14}(i) > 80$: Se o IFR de 14 períodos é maior que 80, então venda, pois o ativo parece estar sobrecomprado.

Tabela 24 – Regras lógicas de compra baseadas em indicadores numéricos da AT do SisPG

Ind	Cod	Regras de Compra
1	RC01	$\text{And}(\text{PrecoAtivo}(i-1) < \text{BBI}_{20}(i-1), \text{PrecoAtivo}(i) > \text{BBI}_{20}(i))$
2	RC02	$\text{And}(\text{PrecoAtivo}(i-1) < \text{BBI}_{14}(i-1), \text{PrecoAtivo}(i) > \text{BBI}_{14}(i))$
3	RC03	$\text{And}(\text{PrecoAtivo}(i-1) < \text{BBI}_{24}(i-1), \text{PrecoAtivo}(i) > \text{BBI}_{24}(i))$
4	RC04	$(\text{IFR}_2(i, 1) < 10)$
5	RC05	$(\text{IFR}_{14}(i, 1) < 30)$
6	RC06	$(\text{IFR}_{14}(i, 1) < 20)$
7	RC07	$(\text{WR}_{14}(i) < -90)$
8	RC08	$(\text{WR}_{21}(i) < -90)$
9	RC09	$(E_{14} < 20)$
10	RC10	$(E_5 < 20)$
13	RC11	$(\text{VROC}_{14} > 80)$
14	RC12	$(\text{VROC}_{14} > 100)$
13	RC13	$(\text{VROC}_{14} > 120)$
14	RC14	$(\text{VROC}_{14} > 150)$
15	RC15	$\text{And}(\text{AD}(i) > \text{AD}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-1))$
16	RC16	$\text{And}(\text{AD}(i) > \text{AD}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-2))$
17	RC17	$\text{And}(\text{AD}(i) > \text{AD}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-3))$
18	RC18	$\text{And}(\text{OBV}(i) > \text{OBV}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-1))$
19	RC19	$\text{And}(\text{OBV}(i) > \text{OBV}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-2))$
20	RC20	$\text{And}(\text{OBV}(i) > \text{OBV}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-3))$
21	RC21	$\text{And}(\text{OBV}(i) > \text{OBV}(i-1), \text{PrecoAtivo}(i) > \text{PrecoAtivo}(i-4))$
22	RC22	$\text{And}(\text{OC}(i) > 0, \text{OC}(i-1) < 0)$
23	RC23	$\text{And}(\text{OC}(i) > 0, \text{OC}(i-2) < 0)$
24	RC24	$\text{And}(\text{OC}(i) > 0, \text{OC}(i-3) < 0)$
25	RC25	$(\text{VC}(i) < 0)$
26	RC26	$(\text{PROC}(i) < -10)$
27	RC27	$(\text{MME}_5(i) > \text{MME}_5(i-3))$
28	RC28	$(\text{MME}_7(i) > \text{MME}_7(i-3))$
29	RC29	$(\text{MME}_8(i) > \text{MME}_8(i-3))$
30	RC30	$(\text{MME}_9(i) > \text{MME}_9(i-3))$
31	RC31	$(\text{MME}_{21}(i) > \text{MME}_{21}(i-3))$
32	RC32	$(\text{MMS}_5(i) > \text{MMS}_5(i-3))$
33	RC33	$(\text{MMS}_7(i) > \text{MMS}_7(i-3))$
34	RC34	$(\text{MMS}_8(i) > \text{MMS}_8(i-3))$
35	RC35	$(\text{MMS}_9(i) > \text{MMS}_9(i-3))$
36	RC36	$(\text{MMS}_{21}(i) > \text{MMS}_{21}(i-3))$

Fonte: Elaborada pelo autor

Tabela 25 – Regras lógicas de venda baseadas em indicadores numéricos da AT do SisPG

Ind	Cod	Regras de Venda
1	RV01	$\text{And}(\text{PrecoAtivo}(i-1) < BBS_{20}(i-1), \text{PrecoAtivo}(i) > BBS_{20}(i))$
2	RV02	$\text{And}(\text{PrecoAtivo}(i-1) < BBS_{14}(i-1), \text{PrecoAtivo}(i) > BBS_{14}(i))$
3	RV03	$\text{And}(\text{PrecoAtivo}(i-1) < BBS_{24}(i-1), \text{PrecoAtivo}(i) > BBS_{24}(i))$
4	RV04	$(IFR_2(i, 1) > 95)$
5	RV05	$(IFR_{14}(i, 1) > 70)$
6	RV06	$(IFR_{14}(i, 1) > 80)$
7	RV07	$(WR_{14}(i) > -10)$
8	RV08	$(WR_{21}(i) > -10)$
9	RV09	$(E_{14} > 90)$
10	RV10	$(E_5 > 90)$
11	RV11	$(VROC_{14} > 80)$
12	RV12	$(VROC_{14} > 100)$
13	RV13	$(VROC_{14} > 120)$
14	RV14	$(VROC_{14} > 150)$
15	RV15	$\text{And}(AD(i) < AD(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-1))$
16	RV16	$\text{And}(AD(i) < AD(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-2))$
17	RV17	$\text{And}(AD(i) < AD(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-3))$
18	RV18	$\text{And}(OBV(i) < OBV(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-1))$
19	RV19	$\text{And}(OBV(i) < OBV(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-2))$
20	RV20	$\text{And}(OBV(i) < OBV(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-3))$
21	RV21	$\text{And}(OBV(i) < OBV(i-1), \text{PrecoAtivo}(i) < \text{PrecoAtivo}(i-4))$
22	RV22	$\text{And}(OC(i) < 0, OC(i-1) > 0)$
23	RV23	$\text{And}(OC(i) < 0, OC(i-2) > 0)$
24	RV24	$\text{And}(OC(i) < 0, OC(i-3) > 0)$
25	RV25	$(VC(i) > 100)$
26	RV26	$(PROC(i) > 10)$
27	RV27	$(MME_5(i) < MME_5(i-3))$
28	RV28	$(MME_7(i) < MME_7(i-3))$
29	RV29	$(MME_8(i) < MME_8(i-3))$
30	RV30	$(MME_9(i) < MME_9(i-3))$
31	RV31	$(MME_{21}(i) < MME_{21}(i-3))$
32	RV32	$(MMS_5(i) < MMS_5(i-3))$
33	RV33	$(MMS_7(i) < MMS_7(i-3))$
34	RV34	$(MMS_8(i) < MMS_8(i-3))$
35	RV35	$(MMS_9(i) < MMS_9(i-3))$
36	RV36	$(MMS_{21}(i) < MMS_{21}(i-3))$

Fonte: Elaborada pelo autor

A partir das tabelas 24 e 25, é possível observar que são 36 regras de compra e 36 regras de venda distintas. O uso do conjunto total de regras implicaria em um grande espaço de busca, o que poderia retardar significativamente a convergência da PG Multiobjetivo.

A Seleção de Características tem como objetivo restringir o número de atributos de entrada para a etapa da PG. A finalidade é reduzir a cardinalidade do espaço de pesquisa a ser explorado. Esse procedimento é feito selecionando as N_a regras mais rentáveis, apresentadas nas tabelas 24 e 25. O procedimento é apresentado no algoritmo 13: este procedimento verifica a rentabilidade de cada par de regras apresentado nas tabelas 24 e 25 em uma determinada janela de tempo. As N_a regras mais lucrativas são selecionadas para a próxima etapa.

Algorithm 13 Pseudocódigo da Seleção de Características

```

function FEATURESELECTION(SlidingWindowI, SetAllRules,  $N_a$ )
     $L \leftarrow \text{length}(\text{SetAllRules})$ ;
    for  $n = 1$  to  $L$  do
        SetRule  $\leftarrow$  Account(SetAllRules( $n$ ), SlidingWindowI);
    end for
    SetRule  $\leftarrow$  Sort(SubSetRule, 'descendent');
    SetRule  $\leftarrow$  SelectRule(SetRule,  $1 : N_a$ );
    return (SetRule);
end function

```

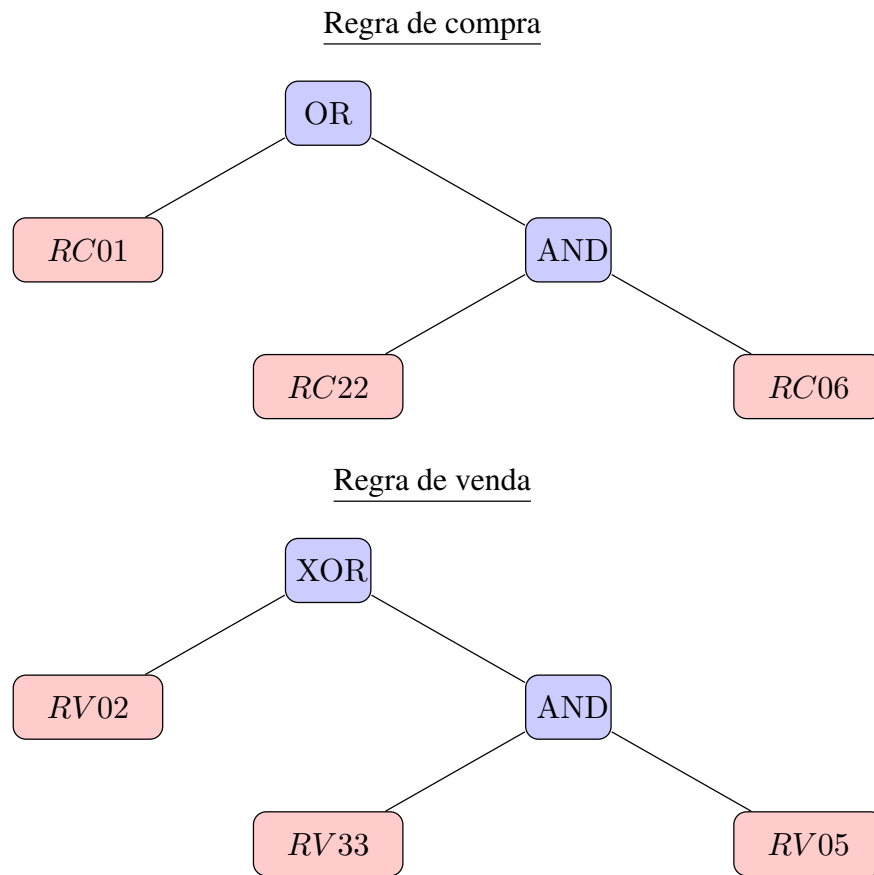
Onde:

- *SlidingWindowI*: é a série financeira de onde se pretende testar as regras apresentadas nas tabelas 24 e 25 e selecionar as N_a melhores regras de negociação;
- *SetAllRules*: conjunto total de regras de compra e de venda apresentadas nas tabelas 24 e 25;
- N_a : número de regras a serem selecionadas, parâmetro definido pelo usuário;
- **Account**: este algoritmo retorna a rentabilidade (lucro ou prejuízo) que uma dada regra obteve, sendo submetida a uma dada janela de tempo em uma dada série financeira;
- **SelectRule**: seleciona as N_a melhores regras.

4.3.5 Programação Genética do SisPG

As regras de investimento do SisPG são geradas usando um algoritmo de PG multiobjetivo. Nesse algoritmo, cada indivíduo codifica duas regras lógicas, uma para compra e outra para venda. Essas regras são formadas pelos atributos (regras da AT) que sobreviveram após a fase de Seleção de Características e os operadores lógicos *AND*, *OR* e *XOR*. Um exemplo de um indivíduo candidato é mostrado na figura 31.

Figura 31 – Exemplo de indivíduo



Fonte: Elaborada pelo autor

É importante destacar que este esquema de codificação é muito interessante, porque torna a implementação de operadores de cruzamento e mutação simples: uma vez que todos os nós são lógicos, qualquer operação de substituição de nó leva a uma solução válida. Logo, não é necessário lidar com o problema de infactibilidade, comum na PG.

Cada indivíduo da população deve tomar uma de três decisões possíveis de cada vez: (i) comprar a ação; (ii) vender a ação ou (iii) manter o estado atual. Essa decisão é tomada com base nos sinais emitidos pelas regras de compra e de venda, como mostrado na tabela 26.

Tabela 26 – Tomada de decisão individual

Regra de Compra	Regra de Venda	Decisão
Falso	Falso	Manter
Verdadeiro	Falso	Comprar
Falso	Verdadeiro	Vender
Verdadeiro	Verdadeiro	Manter

Fonte: Elaborada pelo autor

A geração de regras na etapa de treinamento foi modelada como sendo um problema

biobjetivo, cujos objetivos são: (i) maximizar o retorno financeiro (R) e (ii) reduzir a complexidade do classificador (C). A primeira função é simplesmente o lucro acumulado na janela de treinamento, como apresentado na equação 4.3. Nessa equação, $P(d)$ é o retorno financeiro diário d ($P(d) > 0$ indica lucro e $P(d) < 0$ indica prejuízo).

$$R = \sum_{d=t-\Delta_T}^{t-1} P(d) \quad (4.3)$$

O segundo objetivo destina-se a reduzir a probabilidade de *overfitting*. De acordo com o *Bias-Variance Dilemma*, os classificadores mais complexos têm mais flexibilidade estrutural para modelar, com precisão, os dados de treinamento. Isso pode levar a pequenos erros no conjunto de treinamento, mas baixa capacidade de generalização (ABBASS, 2001). Atualmente, o controle da complexidade de regressores e classificadores é uma prática usual. No algoritmo proposto, a complexidade foi medida como a profundidade máxima¹ nas árvores que representam as regras de compra e de venda. Esse objetivo é mostrado na equação 4.4. Nela, o D_B e D_S representam, a profundidade das regras de compra e de venda, respectivamente.

$$C = \max(D_B, D_S) \quad (4.4)$$

A PG Multiobjetivo foi implementada usando o método *Grow* para gerar soluções iniciais, o cruzamento por ponto e a mutação *Shrink* (KOZA, 1992). Foi utilizado o algoritmo *Fast Non-dominated Sorting* com o método de seleção torneio estocástico, tal como proposto por Deb et al. (2002) e descrito em detalhes no capítulo 3, seção 3.5. O algoritmo de treinamento entrega para o Comitê de Decisão as soluções (indivíduos) não dominadas obtidas.

Na tabela 27, é apresentada a configuração completa do algoritmo de PG Multiobjetivo e no algoritmo 14 é detalhado o procedimento.

Algorithm 14 Pseudocódigo do Treinamento

```

function TRAINING(TrainData, SubSetExpression)
    PopEnsemble  $\leftarrow$  NSGAII(TrainData, SubSetExpression,  $Fo_1$ ,  $Fo_2$ );
    return (PopEnsemble);
end function

```

Onde:

- *TrainData*: é a série financeira utilizada para gerar as regras de compra e venda;
- *SubSetExpressions*: conjunto de regras selecionadas da fase Seleção de Características;
- **NSGAII**: algoritmo evolucionário multiobjetivo proposto por Deb et al. (2002) e descrito em detalhes no capítulo 3;

¹ A profundidade de uma árvore é o comprimento do caminho mais longo entre a raiz e as folhas (CORMEN et al., 2009).

Tabela 27 – Configuração da PG Multiobjetivo

Construção da Árvore	<i>Grow</i>
Método de Seleção	Torneio Estocástico
Método de Cruzamento	Por ponto
Método de Mutação	<i>Shrink</i>
Taxa de Cruzamento	90%
Taxa de Mutação	40%
Tamanho da População	60
Função Objetivo 1	$F_{o1} = \text{Maximizar}(R)$
Função Objetivo 2	$F_{o2} = \text{Minimizar}(C)$
Número de Gerações	60

Fonte: Elaborada pelo autor

- F_{o1} : *Maximizar*(R) 4.3;
- F_{o2} : *Minimizar*(C) 4.4.

4.3.6 Comitê de Decisão do SisPG

A composição do Comitê de Decisão é feita pelas soluções (indivíduos) não dominadas geradas na fase de treinamento. Já a decisão de comprar ou de vender uma ação é feita por um sistema de votação da maioria. Cada árvore de decisão de compra e de venda de cada solução é avaliada, dia a dia. Se a maioria indicar uma compra, é retornado um sinal de compra. Se a maioria indicar uma venda, é retornado um sinal de venda. O algoritmo do Comitê de Decisão do SisPG é o mesmo já apresentado no SisSM (ver algoritmo 9).

4.3.7 Sistema de negociação do SisPG

O sistema de negociação do SisPG é o mesmo do SisSM apresentado na subseção 4.2.6.

5 Resultados

Neste capítulo, são apresentados a estratégia dos testes e os resultados dos experimentos realizados com o Sistema Automatizado de Investimento via Seleção Multiobjetivo de Regras (SisSM) e o Sistema Automatizado de Investimento via Programação Genética Multiobjetivo (SisPG). Os testes foram realizados com seis ações da BOVESPA¹, a saber: BBAS3, BOVA11, CMIG4, EMBR3, GGBR4, VALE5. Essas ações foram selecionadas pelos seguintes critérios:

- Possuem representatividade no índice BOVESPA e apresentam liquidez para as operações de compra e de venda em *Candles* diários;
- São negociadas também na Bolsa de Valores de Nova York, através das “American Depositary Receipts”(ADRs), o que aumenta a liquidez do ativo;
- Estão distribuídas em quatro setores distintos da economia brasileira: Comodite, Energia, Financeiro e Indústria;
- As séries apresentam um comportamento diversificado no período dos testes: BOVA11, GGBR4 e VALE5 em tendência de baixa. BBAS3 e CMIG4 apresentaram lateralização, e EMBR3 em tendência de alta.

Na tabela 28, é apresentada uma breve descrição de cada uma dessas ações.

Tabela 28 – Código e descrição das ações utilizadas nos testes do SisSM e SisPG

Ação	Empresa	Setor	Tendência	Site
BBAS3	Banco do Brasil S.A	Financeiro	Lateralização	www.bb.com.br
BOVA11	ETF	80% Bovespa	Baixa	www.bmfbovespa.com.br
CMIG4	CEMIG S.A	Energia	Lateralização	www.cemig.com.br
EMBR3	EMBRAER S.A	Indústria	Alta	www.embraer.com
GBR4	GERDAU S.A	Indústria	Baixa	www.gerdau.com
VALE5	VALE S.A	Comodite	Baixa	www.vale.com

Fonte: Elaborada pelo autor

¹ Índice Bovespa (Ibovespa) é o mais importante indicador do desempenho médio das cotações das ações negociadas na Bolsa de Valores de São Paulo. É formado pelas ações com maior volume negociado nos últimos meses. O valor atual representa a quantia, em moeda corrente, de uma carteira teórica de ações, constituída em 2 de janeiro de 1968, a partir de uma aplicação hipotética. (WIKIPEDIA, 2017a)

5.1 Estratégias de Testes do SisSM

A tabela 29 indica os parâmetros que foram utilizados nos testes do SisSM. Eles foram escolhidos com base em experimentos preliminares, realizados em janelas de teste antes de fevereiro de 2013.

Tabela 29 – Parâmetros adotados nos testes do SisSM

Parâmetros	Valores
t_0	05.02.2013
t_{final}	02.02.2015
Δ_F (dias)	30
Δ_T (dias)	30
Δ_E (dias)	10
N_a	50
N_b	5
$MaxLoss$	6%
$MaxGain$	12%

Fonte: Elaborada pelo autor

O SisSM foi comparado com outras duas abordagens:

Buy and Hold (B&H): Esta é uma estratégia passiva de investimento em que um investidor compra as ações e as mantém por um longo período de tempo, independentemente das flutuações no mercado. O investidor que emprega a estratégia *Buy and Hold* não se preocupa com movimentos de preços de curto prazo e nem com os indicadores da AT;

Comitê de Decisão com todas as regras da AT (AT-34): Esta estratégia também utiliza janelas móveis dentro do laço de repetição, o mesmo Comitê de Decisão e o mesmo sistema de negociação propostos no trabalho. A diferença é que não existe seleção de regras, ou seja, todas as 68 regras apresentadas nas tabelas 4 e 5 são utilizadas no comitê de decisão, para as decisões de compra e de venda.

5.2 Resultados do SisSM

O retorno financeiro obtido pelo SisSM juntamente com as duas abordagens apresentadas na seção 5.1 são mostrados na tabela 30. Todos os valores são relatados como valores únicos porque os sistemas são determinísticos (isto é, obtêm-se sempre os mesmos resultados quando aplicados à mesma janela de tempo).

Os resultados apresentados na tabela 30 indicam que o SisSM obteve um retorno superior às outras duas estratégias. O SisSM também conseguiu lucro com as séries financeiras das ações nas três tendências (alta, baixa e lateralizado), com destaque para as ações de BBAS3, onde o lucro atingiu 231.78% de retorno.

Tabela 30 – Retorno financeiro obtido pelo SisSM e as duas estratégias apresentadas

Ação	SisSM	B&H	AT-34
BBAS3	231.78%	0.60%	72.40%
BOVA11	42.02%	-20.44%	37.25%
CMIG4	84.81%	10.83%	3.37%
EMBR3	107.54%	46.57%	106.33%
GGBR4	139.78%	-44.37%	15.97%
VALE5	48.42%	-48.75%	31.47%
Portfólio	109,06%	-9.26%	44,47%

Fonte: Elaborada pelo autor

Após serem realizados os experimentos apresentados na tabela 30 e observados os resultados de TA-34, optou-se por investigar a qualidade do par de regras de compra e de venda apresentado nas tabelas 4 e 5. Não foram exploradas as 1.156 combinações possíveis, mas os pares (RC_i RV_i) apresentados nas tabelas 4 e 5, para (i) de 1 a 34. Os experimentos foram feitos utilizando os mesmos parâmetros expostos na tabela 29 e o mesmo módulo de negociação utilizado pelo SisSM. Na tabela 31 são apresentados os resultados.

Nesse experimento, de fato verificou-se que alguns pares de regras obtiveram retornos individuais superiores ao retorno financeiro do SisSM. Na tabela 31 esses retornos estão destacados em negrito. Para BOVA11, os pares (RC10,RV10), (RC19,RV19), (RC23,RV23), (RC24,RV24), (RC25,RV25); para EMBR3, o par (RC33,RV33) e, finalmente, para VALE5, o par (RC32,RV32), os resultados foram superiores aos retornos conseguidos pelo SisSM, o que não invalida os méritos obtidos pelo sistema proposto, uma vez que ele foi capaz de selecionar, entre 1.156 combinações, as melhores para compor o Comitê de Decisão e efetuar as operações de compra e de venda, quando necessário.

Também é importante destacar que os pares que obtêm resultados superiores em alguns casos perdem em vários outros. Ou seja, não tem uma regra que seja globalmente a melhor, mas o SisSM neste experimento sempre conseguiu escolher um conjunto de regras adequado.

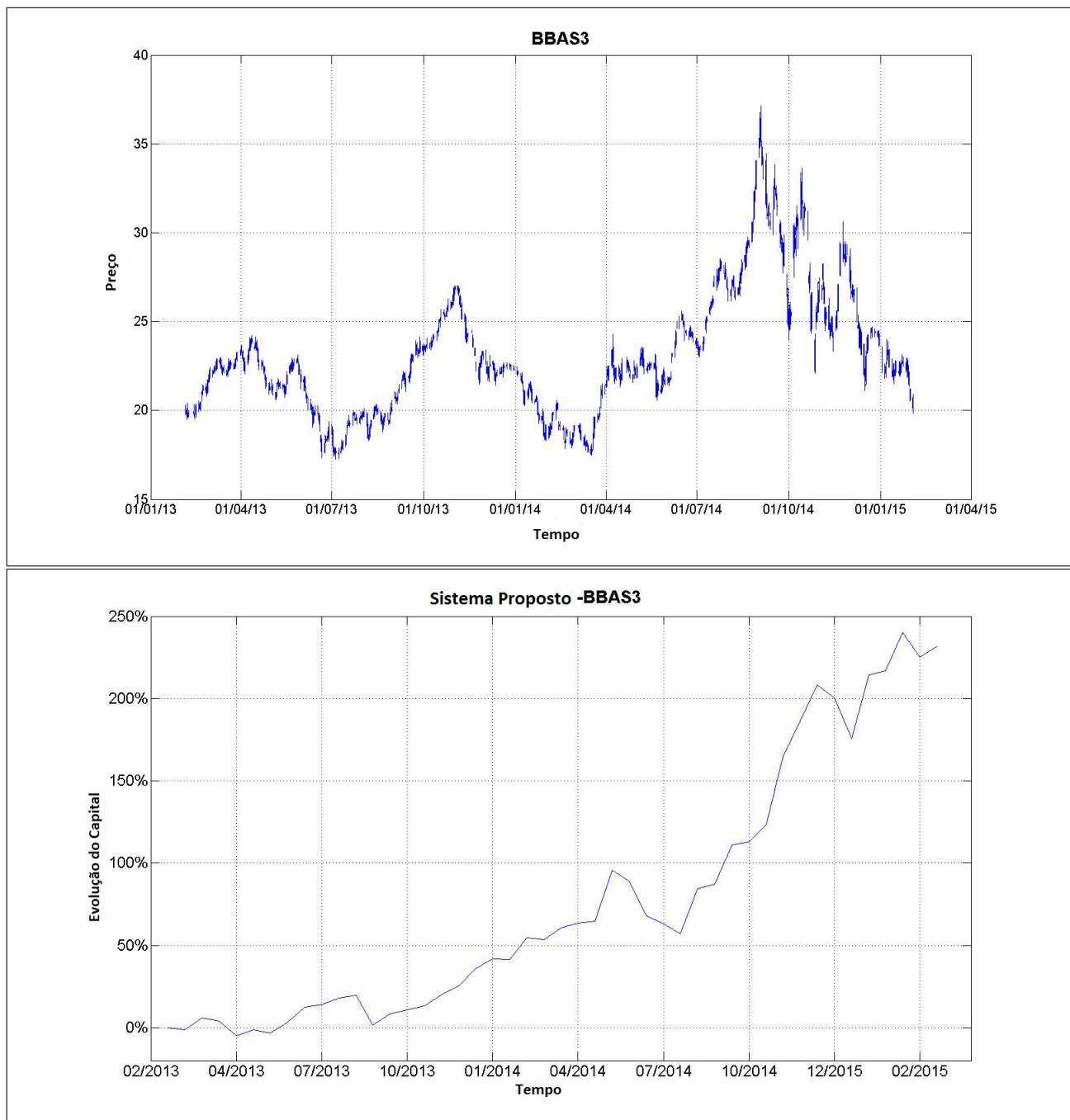
Nessa seção foram comparados o desempenho do sistema proposto com o retorno da ação no período, como pode ser observado nas figuras 32, 33, 34, 35, 36 e 37.

Tabela 31 – Retorno financeiro obtido pelo par de regras da AT utilizadas no SisSM

	RCI	RVi	BBAS3 (%)	BOVA11 (%)	CMIG4 (%)	EMBR3 (%)	GGBR4 (%)	VALE5 (%)
1	RC01	RV01	-7,34	16,13	-14,96	3,20	39,14	13,10
2	RC02	RV02	-15,96	-3,00	-10,30	3,88	38,01	-1,38
3	RC03	RV03	10,01	21,30	-34,95	18,39	23,37	17,12
4	RC04	RV04	-113,99	-6,85	15,23	0,83	70,02	-11,86
5	RC05	RV05	1,80	23,18	20,94	62,49	63,13	40,84
6	RC06	RV06	0,60	9,98	13,06	24,75	21,44	37,50
7	RC07	RV07	-1,41	17,07	53,88	70,35	43,47	11,06
8	RC08	RV08	16,41	18,60	26,60	57,54	32,25	8,66
9	RC09	RV09	23,00	31,62	66,98	65,75	38,95	20,67
10	RC10	RV10	30,34	45,34	14,29	26,16	20,18	34,58
11	RC11	RV11	-121,92	-4,48	20,45	9,38	58,93	-17,16
12	RC12	RV12	-109,68	-6,85	20,34	1,40	71,47	-11,86
13	RC13	RV13	0,00	0,00	0,00	0,00	0,00	0,00
14	RC14	RV14	0,00	0,00	0,00	0,00	0,00	0,00
15	RC15	RV15	87,85	10,05	2,65	-6,61	-12,65	-24,76
16	RC16	RV16	0,00	0,00	0,00	0,00	0,00	0,00
17	RC17	RV17	0,00	0,00	0,00	0,00	0,00	0,00
18	RC18	RV18	64,78	-7,66	45,25	-19,85	18,82	-1,26
19	RC19	RV19	185,37	47,00	15,08	12,91	-39,21	-6,19
20	RC20	RV20	105,55	17,06	-39,84	-6,97	-95,82	-5,68
21	RC21	RV21	71,52	38,53	-65,25	-15,74	-111,46	3,60
22	RC22	RV22	60,49	41,69	-43,60	-13,42	-33,72	25,15
23	RC23	RV23	17,87	46,05	-51,09	-11,11	-42,86	39,77
24	RC24	RV24	5,21	42,64	-52,47	-13,90	-47,01	32,65
25	RC25	RV25	54,31	25,09	58,09	1,82	-44,85	-17,25
26	RC26	RV26	40,29	18,23	26,88	7,29	42,89	24,03
27	RC27	RV27	-74,94	9,09	4,95	53,94	46,30	31,35
28	RC28	RV28	8,15	16,79	38,45	47,69	49,97	19,14
29	RC29	RV29	40,32	16,81	-1,08	98,64	68,41	2,54
30	RC30	RV30	28,94	34,19	10,02	43,15	20,80	7,23
31	RC31	RV31	-130,84	-10,91	-30,12	4,23	33,00	0,12
32	RC32	RV32	61,38	47,61	73,56	19,96	136,17	100,33
33	RC33	RV33	75,65	31,88	68,34	110,80	53,07	-6,76
34	RC34	RV34	62,01	-20,34	-28,87	-1,59	-23,36	-0,96

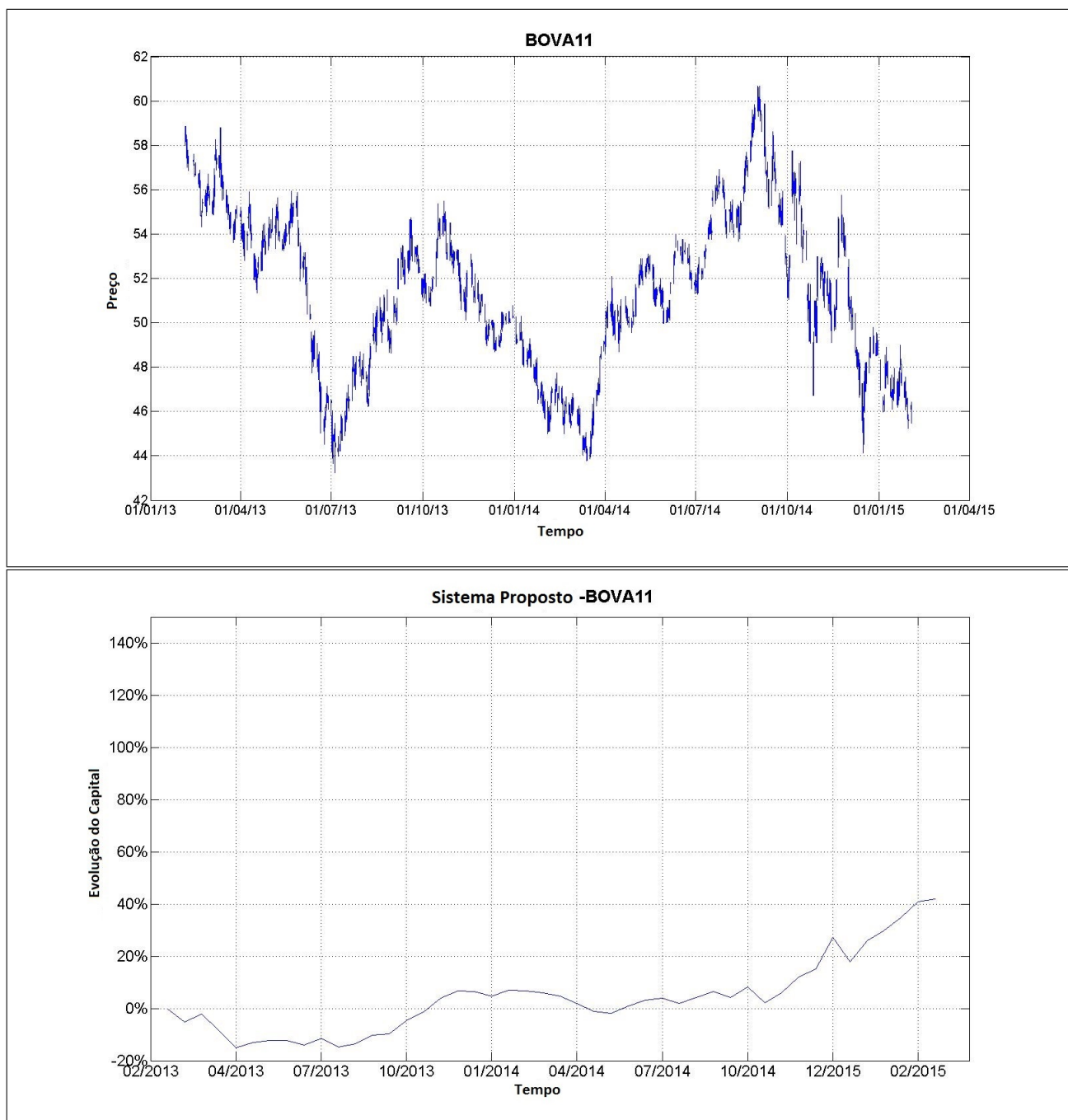
Fonte: Elaborada pelo autor

Figura 32 – BBAS3 - Valor da Ação X Desempenho do SisSM



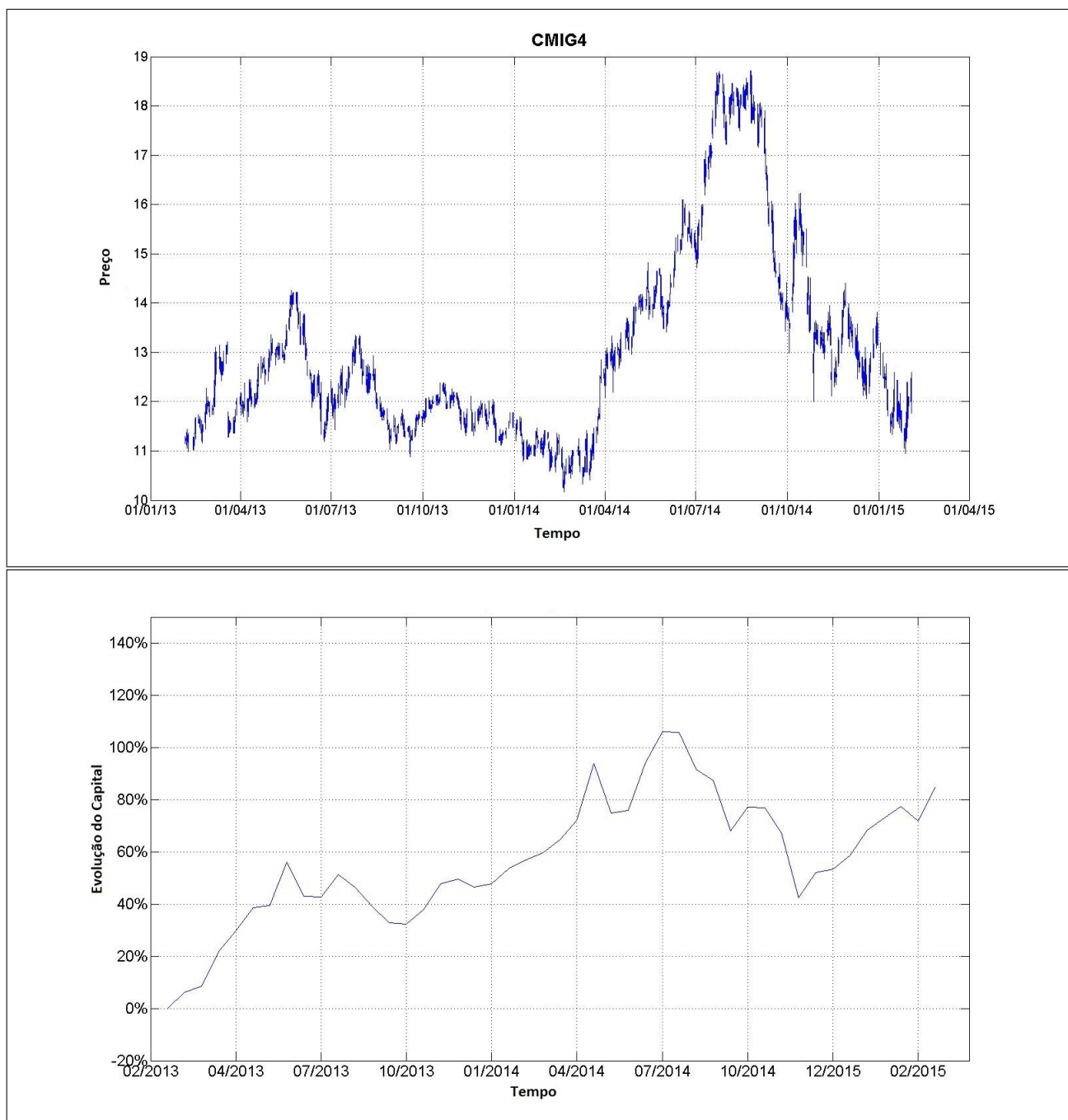
Fonte: Elaborada pelo autor

Figura 33 – BOVA11 - Valor da Ação X Desempenho do SisSM



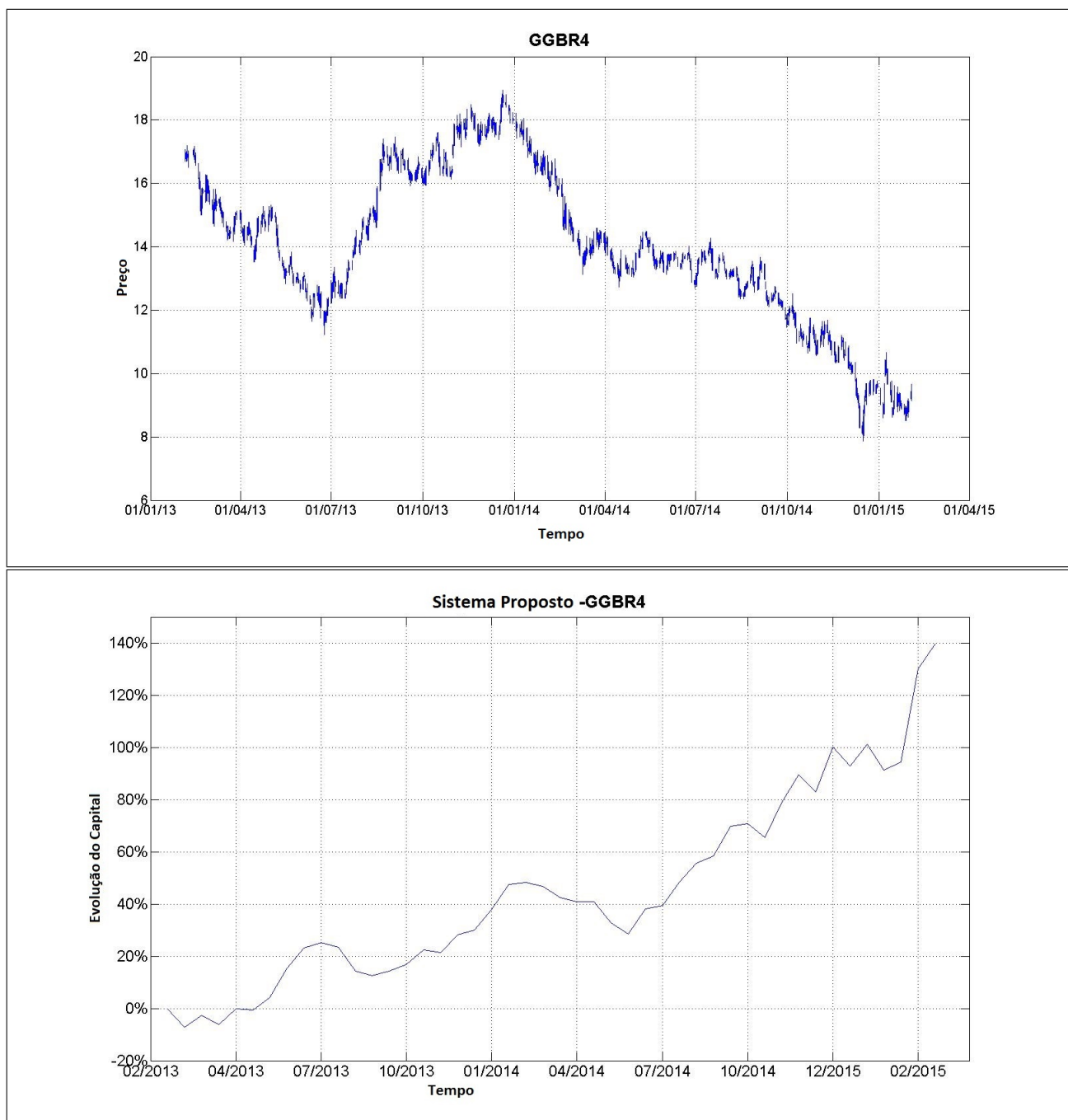
Fonte: Elaborada pelo autor

Figura 34 – CMIG4 - Valor da Ação X Desempenho do SisSM



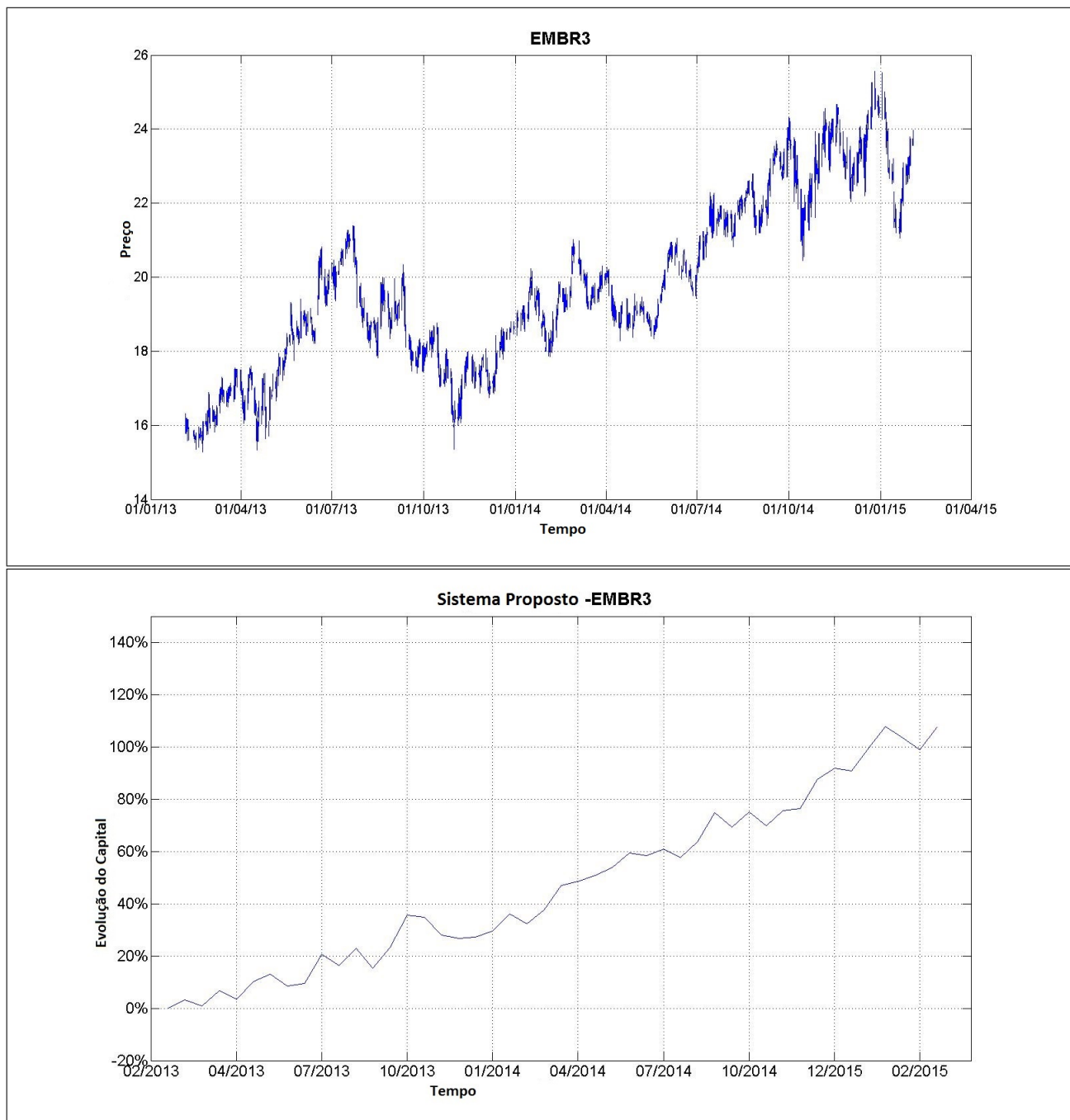
Fonte: Elaborada pelo autor

Figura 35 – GGBR4 - Valor da Ação X Desempenho do SisSM



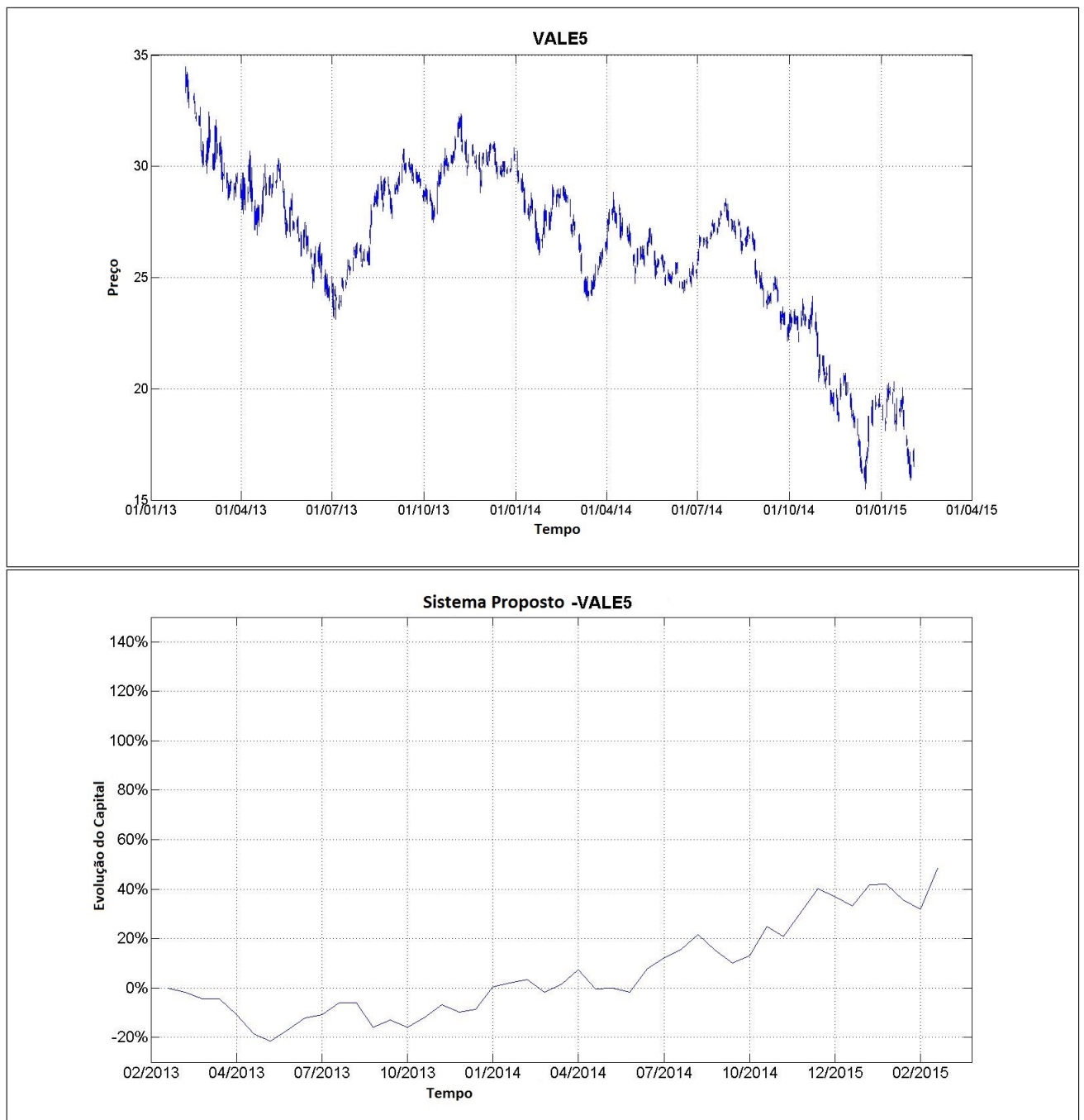
Fonte: Elaborada pelo autor

Figura 36 – EMBR3- Valor da Ação X Desempenho do SisSM



Fonte: Elaborada pelo autor

Figura 37 – VALE5 - Valor da Ação X Desempenho do SisSM



Fonte: Elaborada pelo autor

5.2.1 Teste de Estresse no SisSM

Um segundo experimento foi realizado com o objetivo de avaliar o SisSM em situações de crise. O experimento foi realizado entre 07/07/2015 e 14/07/2016 (260 dias úteis). Nesse período, o Brasil passou por cenários econômicos e políticos muito turbulentos. Em termos econômicos, a moeda brasileira (Real) perdeu cerca de 50% de seu valor em relação ao dólar, ao se comparar com o ano de 2014. No cenário político, a recém-eleita presidente do Brasil Dilma Rousseff, sofreu um longo processo de *impeachment*, sendo deposta do governo brasileiro em agosto de 2016 (WIKIPEDIA, 2016). Obviamente, esse cenário impôs uma forte oscilação ao mercado financeiro, com uma tendência de queda global nas ações da BOVESPA.

Os resultados obtidos pelo sistema proposto, comparados com as outras duas abordagens, são apresentados na Tabela 32. Eles sugerem que o SisSM não foi adequado neste caso.

Tabela 32 – Retorno financeiro obtido pelo SisSM e as duas estratégias apresentadas (Período de crise)

Ação	SisSM	B&H	AT-34
BBAS3	44.62%	-18.33%	-29.73%
BOVA11	-20.05%	5.69%	-3.37%
CMIG4	-39.18%	-22.42%	-0.2%
EMBR3	4.36%	-23.62%	-20.03%
GGBR4	22.27%	-3.28%	61.11%
VALE5	-41.8%	-8.78%	-20.04%
Portfólio	-30.23%	-9.26%	-12.19%

Fonte: Elaborada pelo autor

5.3 Estratégias de Testes do SisPG

A tabela 33 indica os parâmetros que foram utilizados nos testes do SisPG. Eles foram escolhidos com base em experimentos preliminares, realizados em janelas de teste antes de fevereiro de 2013.

Tabela 33 – Parâmetros adotados nos testes do SisPG

Parâmetros	Valores
t_0	05.02.2013
t_{final}	02.02.2015
Δ_F (dias)	30
Δ_T (dias)	30
Δ_E (dias)	10
F_{low}	0.3
N_a	10
Tamanho da População	60
Taxa de cruzamento (por par)	0.90
Taxa de mutação (por indivíduo)	0.40
Número de Gerações	60
F_{lev}	100%
$MaxLoss$	6%
$MaxGain$	12%

Fonte: Elaborada pelo autor

O SisPG foi comparado com outras três abordagens:

Buy and Hold (B&H): Esta é uma estratégia passiva de investimento em que um investidor compra as ações e as mantém por um longo período de tempo, independentemente das flutuações no mercado. O investidor que emprega a estratégia *Buy and Hold* não se preocupa com movimentos de preços de curto prazo e nem com os indicadores da AT.

Comitê de Decisão com as melhores regras da AT (AT-10): Esta estratégia também utiliza janelas móveis dentro do laço de repetição, o mesmo módulo de seleção de características, o Comitê de Decisão e o sistema de negociação propostos no trabalho. Em cada iteração, o módulo de seleção de características é empregado para indicar as N_a melhores regras lógicas de compra e de venda no intervalo. Essas regras são usadas no Comitê de Decisão, que toma as decisões na janela de negociação. A janela de treinamento não é considerada nesta abordagem, uma vez que a PG não é empregada.

Comitê de Decisão com todas as regras da AT (AT-36): Esta estratégia é muito similar à estratégia anterior. A única diferença é que não é usada a seleção de características. Portanto, o Comitê de Decisão é formado por todas as 72 regras apresentadas nas tabelas 24 e 25.

5.4 Resultados do SisPG

O retorno financeiro obtido pelo sistema proposto juntamente com as três abordagens apresentadas na seção 5.3 são mostrados na tabela 34. Os valores relatados pelo sistema proposto são expressos usando a média e o desvio padrão. Os desempenhos dos outros métodos são

expressos como valores únicos, porque são determinísticos (isto é, obtêm-se sempre os mesmos resultados quando aplicados à mesma janela de tempo).

Tabela 34 – Retorno financeiro obtido pelo SisPG e as três estratégias apresentadas

Ação	SisPG	B&H	AT-10	AT-36
BBAS3	77.61%±3.16%	0.60%	6.17%	37.77%
BOVA11	33.37%±2.80%	-20.44%	10.03%	8.35%
CMIG4	54.19%±5.78%	10.83%	29.42%	21.09%
EMBR3	71.62%±6.14%	46.57%	13.49%	-38.71%
GGBR4	48.50%±2.38%	-44.37%	0.78%	-5.79%
VALE5	39.88%±4.62%	-48.75%	-22.10%	3.63%
Portfólio	54.20%	-9.26%	6.30%	4.39%

Fonte: Elaborada pelo autor

A partir dos dados da tabela 34, é possível observar que o método proposto superou claramente as outras três abordagens. Os retornos financeiros obtidos pelo sistema automatizado foram sempre positivos e consideravelmente superiores à estratégia (B&H). É importante notar que o método foi capaz de obter lucro significativo, mesmo em situações em que o ativo depreciou consideravelmente, como no caso das ações de BOVA11, GGBR4 e VALE5. Uma comparação direta com a estratégia AT-10 mostra que a PG teve um impacto importante nos resultados: a diferença média entre os métodos varia de 23% a 70%, dependendo da ação. A comparação entre AT-10 e AT-36 sugere que o módulo de seleção de características também tem um impacto positivo nos resultados finais: a estratégia AT-10 é melhor do que a estratégia AT-36 em 4 de 6 ações.

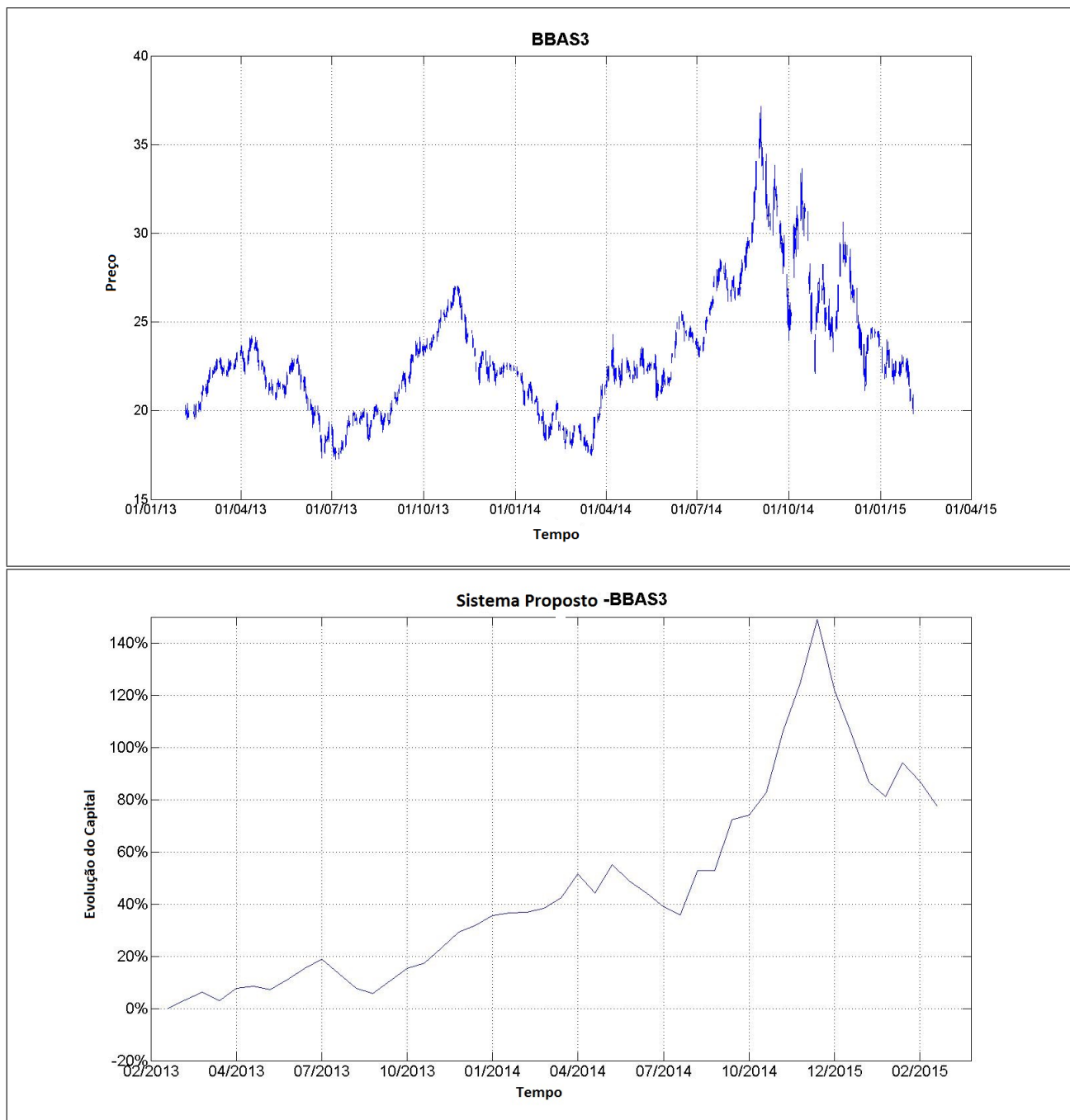
A evolução de uma carteira de investimentos teórica, composta pelas seis ações (com o orçamento inicial dividido igualmente entre as ações), é indicada na última linha da tabela (Portfólio). É possível notar que o SisPG obteve um lucro significativo, enquanto os outros métodos obtiveram rendimentos inferiores a 10%.

A abordagem estatística proposta por Carrano, Wanner e Takahashi (2011) foi aplicada para comparar os retornos financeiros obtidos pelo método proposto e os resultados das três abordagens apresentadas. As hipóteses foram testadas utilizando-se o teste T Student e a significância foi corrigida usando correção de Bonferroni para 18 comparações. Com o nível de confiança de 99%, é possível supor que as diferenças observadas entre a abordagem proposta e os métodos de referência são estatisticamente relevantes, uma vez que o maior valor p observado nas comparações foi inferior a 10^{-4} (a hipótese nula pode ser rejeitada para o valor de p inferior a 5.56×10^{-4}).

Nas figuras 38, 39, 40, 41, 42 e 43 são apresentados a valorização da ação versus o desempenho médio do sistema. Essas figuras são importantes, pois ajudam a compreender os pontos fortes e as fragilidades do sistema ao longo do período testado. No geral, pode-se perceber que em momentos em que a tendência do papel está bem definida, seja na valorização (tendência

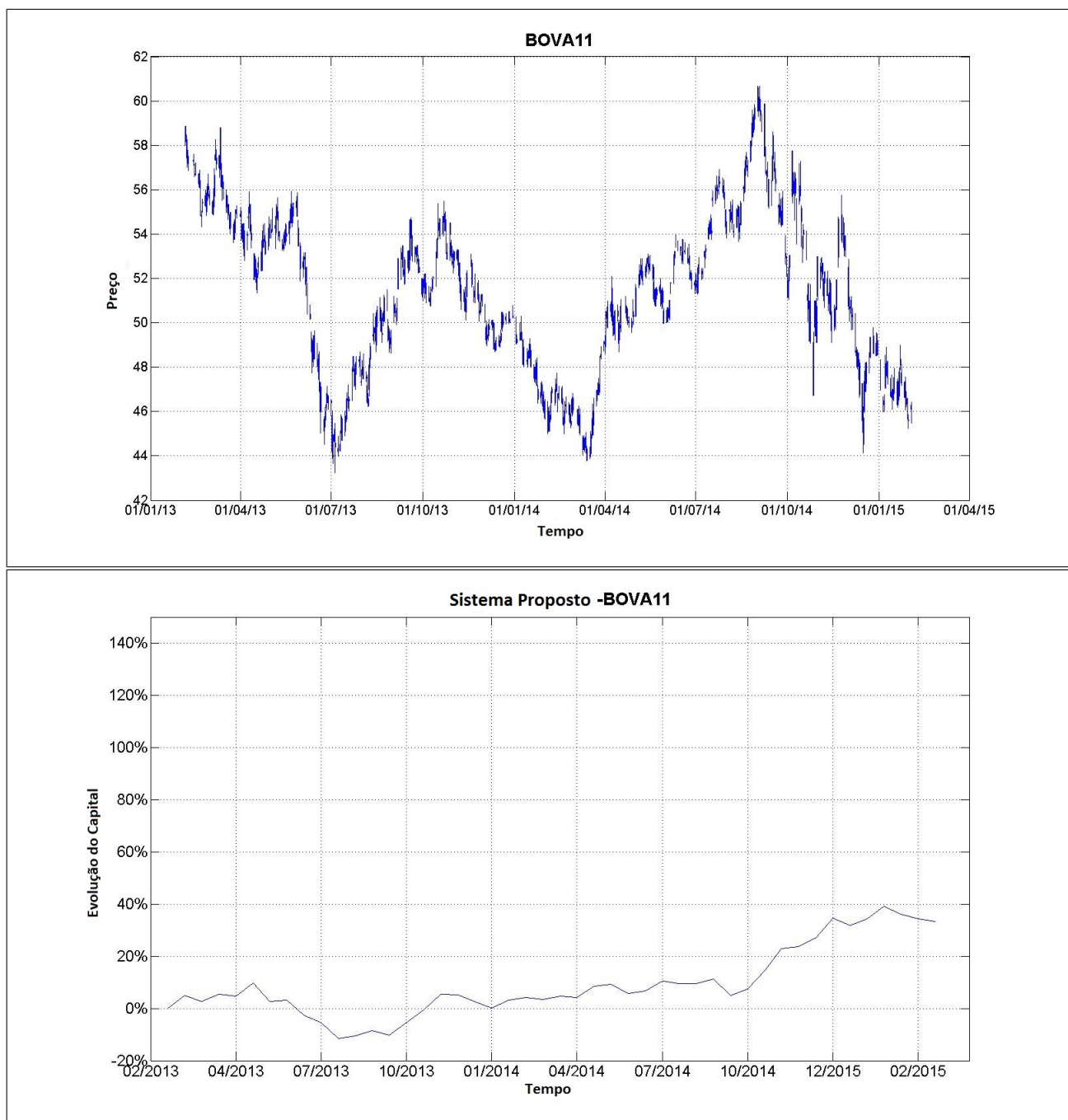
de alta) ou na desvalorização (tendência de baixa), o SisPG apresenta melhor desempenho. O principal ponto fraco é na mudança de tendência. Nessas inversões, seja da alta para a baixa, ou da baixa para a alta, geralmente ocorre redução do capital. Porém, é importante ressaltar que o sistema conseguiu captar a mudança de tendência e retornar a trajetória de lucros.

Figura 38 – BBAS3 - Valor da Ação X Desempenho do SisPG



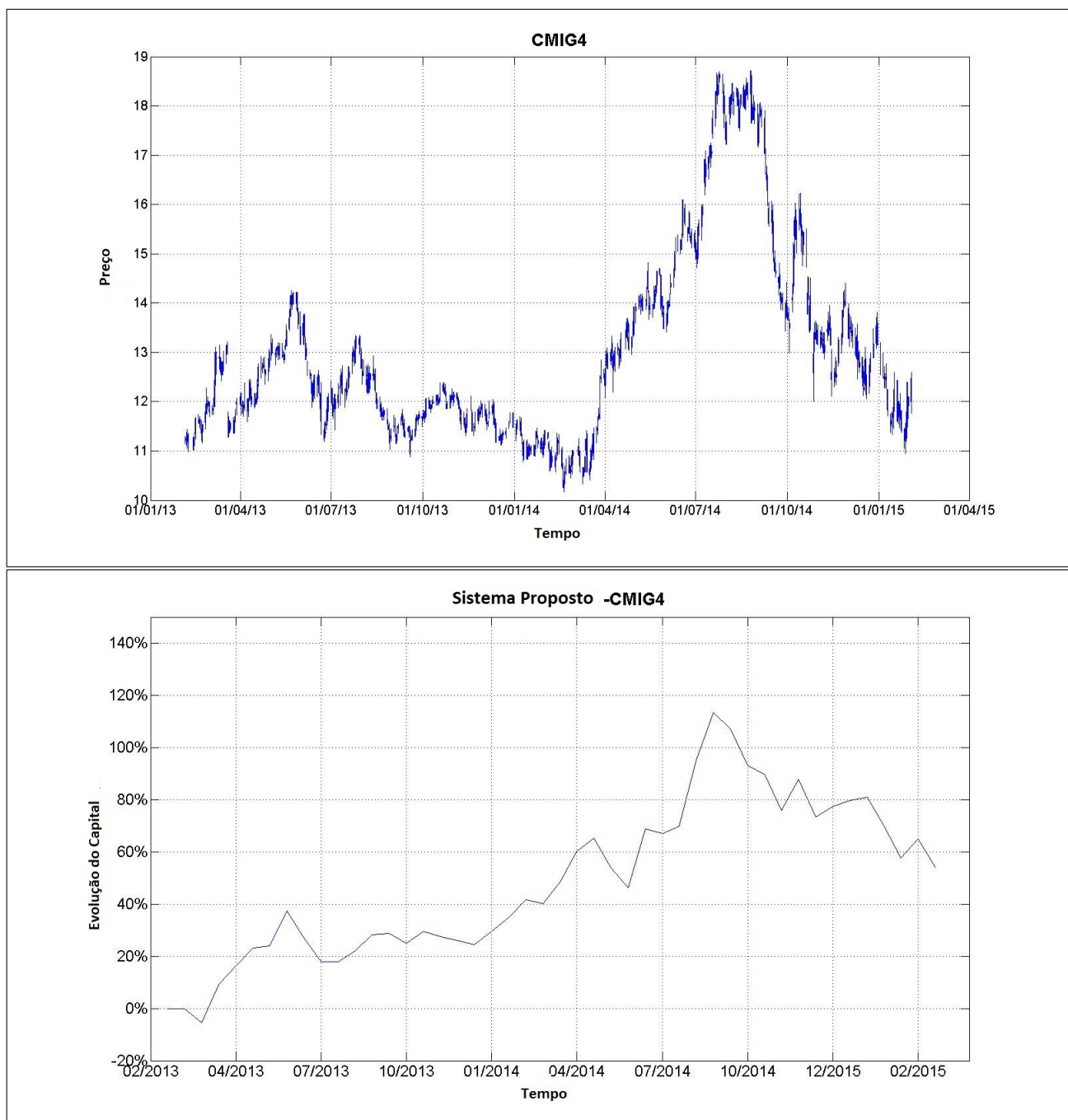
Fonte: Elaborada pelo autor

Figura 39 – BOVA11 - Valor da Ação X Desempenho do SisPG



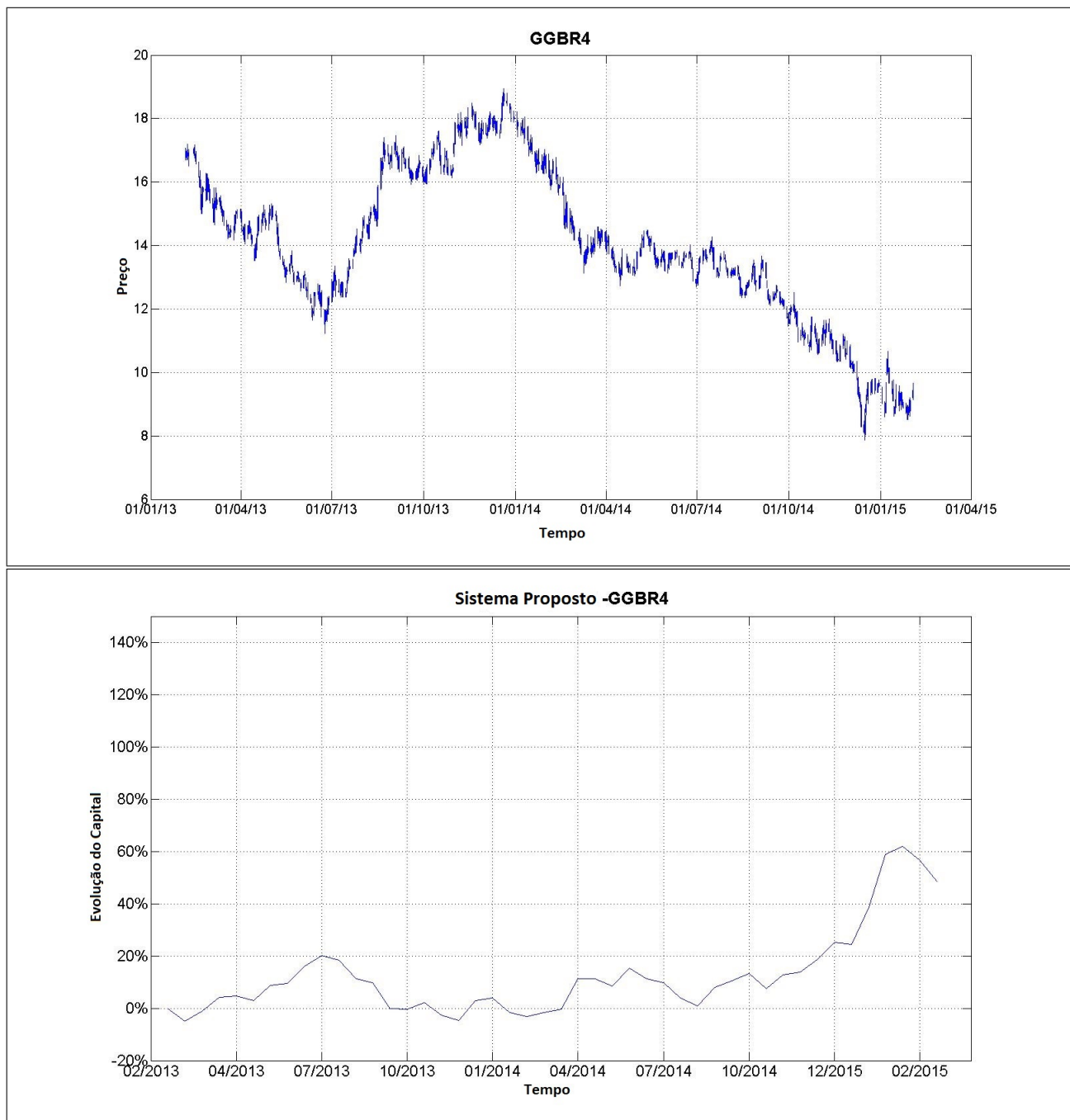
Fonte: Elaborada pelo autor

Figura 40 – CMIG4 - Valor da Ação X Desempenho do SisPG



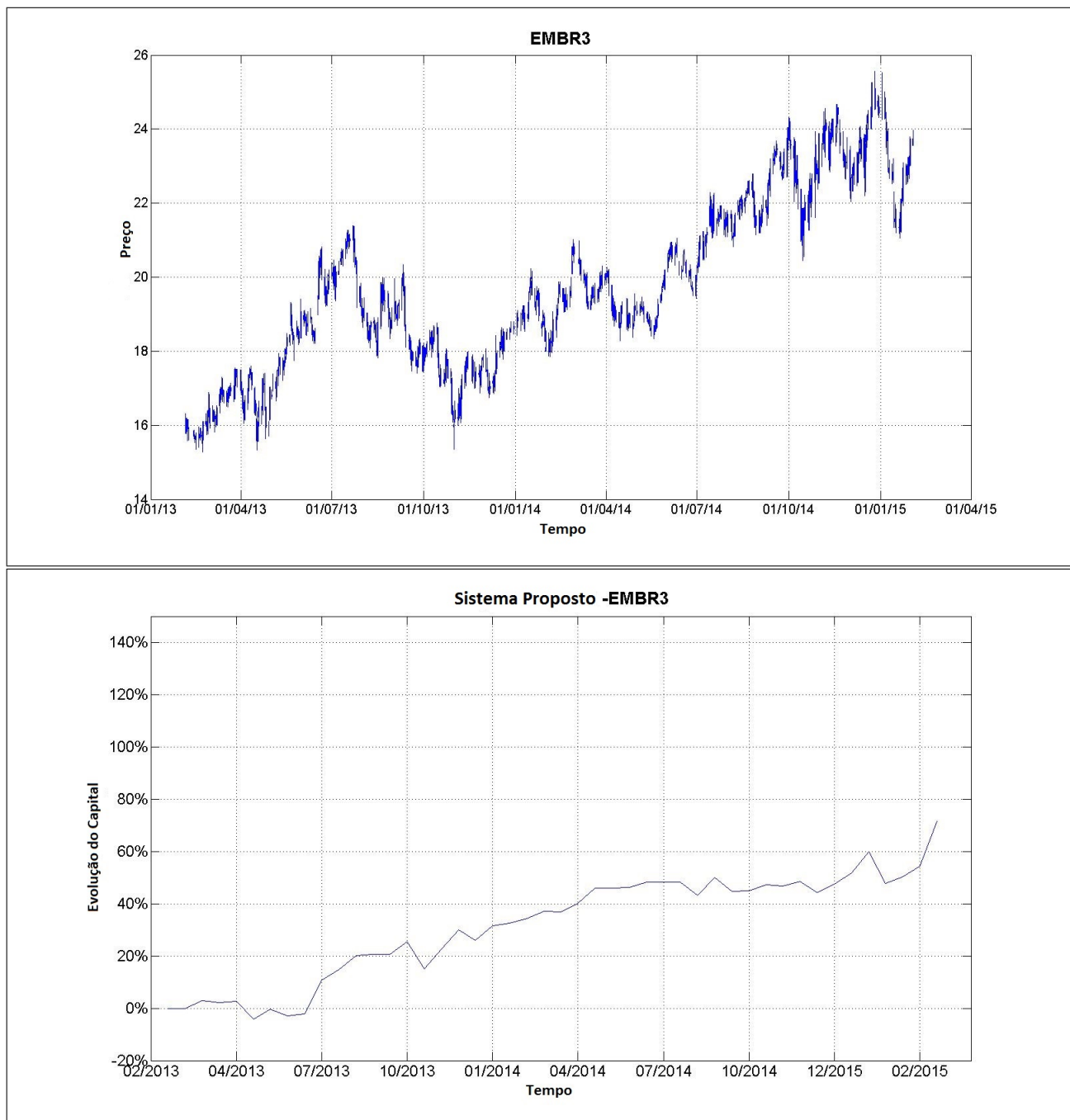
Fonte: Elaborada pelo autor

Figura 41 – GGBR4 - Valor da Ação X Desempenho do SisPG



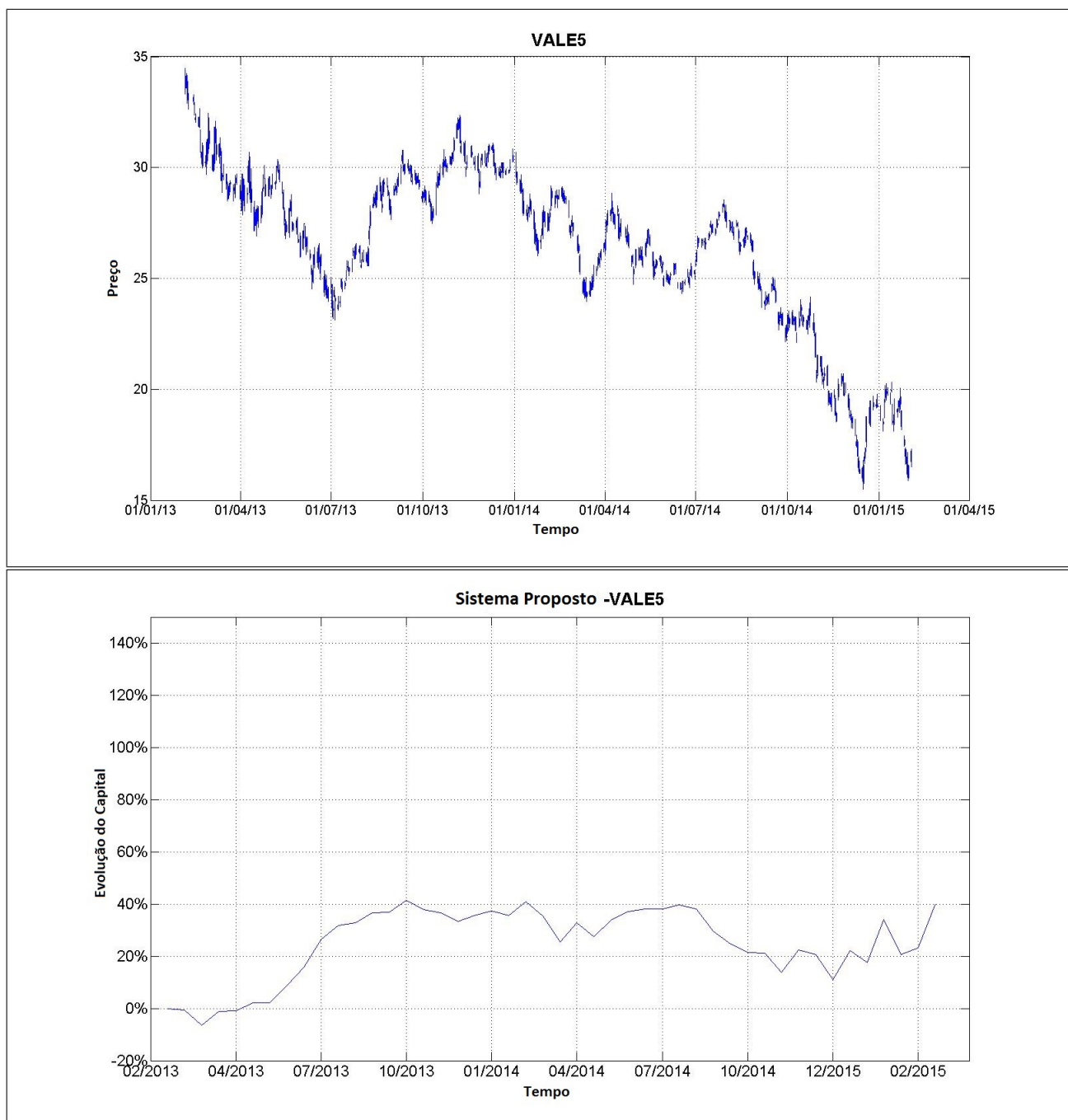
Fonte: Elaborada pelo autor

Figura 42 – EMBR3- Valor da Ação X Desempenho do SisPG



Fonte: Elaborada pelo autor

Figura 43 – VALE5 - Valor da Ação X Desempenho do SisPG



Fonte: Elaborada pelo autor

5.4.1 Teste de Estresse no SisPG

Um segundo experimento também foi realizado no SisPG com objetivo de avaliar o comportamento do sistema em períodos de estresse. O experimento seguiu as mesmas datas adotadas no segundo teste feito com o SisSM, apresentado na subseção 5.4.1. Ele foi feito entre 07/07/2015 e 14/07/2016 (260 dias úteis).

Os resultados obtidos pelo sistema proposto comparados com as outras três abordagens são mostrados na Tabela 35. O algoritmo proposto superou a estratégia (B&H) em cinco comparações, a estratégia (AT-10) em quatro comparações e a estratégia (AT-36) em cinco comparações. Além disso, o retorno médio da carteira de investimento teórica foi consideravelmente maior do que os outros métodos.

Tabela 35 – Retorno financeiro obtido pelo SisPG e as três estratégias apresentadas (Período de Crise)

Ação	SisPG	B&H	AT-10	AT-36
BBAS3	19.83%±2.72%	-18.33%	5.73%	11.13%
BOVA11	3.16%±0.15%	5.69%	9.87%	-1.03%
CMIG4	23.33%±4.61%	-22.42%	8.76%	3.47%
EMBR3	14.36%±1.44%	-23.62%	12.18%	6.43%
GGBR4	-2.11%±1.99%	-3.28%	4.51%	3.92%
VALE5	50,58%±6.18%	-8.78%	2.03%	-3.14%
Portfólio	18.19%	-11.79%	7.18%	3.46%

Fonte: Elaborada pelo autor

5.5 Comparação entre os Resultados do SisPG e SisSM

Após a realização dos dois experimentos apresentados nas tabelas 36 e 37, pode-se observar que, apesar do resultado do SisSM ter sido superior no primeiro experimento, o SisPG apresentou maior capacidade de enfrentar situações de estresse no mercado financeiro. São necessários mais testes mas, com estes resultados, é possível sugerir que o algoritmo evolucionário utilizado no SisPG foi capaz de dar melhor resposta em uma situação de turbulência no mercado financeiro. A possibilidade de gerar novas regras foi determinante para os resultados do segundo experimento.

Tabela 36 – Comparação entre os retornos financeiros obtidos pelo SisPG e os retornos financeiros obtidos pelo SisSM no primeiro experimento

Ação	SisPG	SisSM	B&H
BBAS3	77.61%±3.16%	231.78%	0.60%
BOVA11	33.37%±2.80%	42.02%	-20.44%
CMIG4	54.19%±5.78%	84.81%	10.83%
EMBR3	71.62%±6.14%	107.54%	46.57%
GGBR4	48.50%±2.38%	139.78%	-44.37%
VALE5	39.88%±4.62%	48.42%	-48.75%
Portfólio	54.20%	109.06%	-9.26%

Fonte: Elaborada pelo autor

Tabela 37 – Comparação entre os retornos financeiros obtidos pelo SisPG e os retornos financeiros obtidos pelo SisSM no segundo experimento

Ação	SisPG	SisSM	B&H
BBAS3	19.83%±3.16%	44.62%	-18.33%
BOVA11	3.16%±2.80%	-20.05%	5.69%
CMIG4	23.33%±5.78%	-39.18%	-22.42%
EMBR3	14.36%±6.14%	4.36%	-23.62%
GGBR4	-2.11%±2.38%	22.27%	-3.28%
VALE5	50,58%±4.62%	-41.8%	-8.78%
Portfólio	18.19%	-30.23%	-11.79%

Fonte: Elaborada pelo autor

6 Conclusões e Propostas de Continuidade

6.1 Conclusões

6.1.1 Sistema Automatizado de Investimento via Seleção Multiobjetivo de Regras (SisSM)

A seleção de regras, considerando o desempenho passado e utilizando a otimização multiobjetivo, se mostrou promissora. Porém, os resultados negativos observados em uma situação de crise não podem ser ignorados. O SisSM, sistema desenvolvido neste trabalho, foi testado com as mesmas ações e nas mesmas janelas de testes utilizadas no SisPG.

6.1.2 Sistema Automatizado de Investimento via Programação Genética Multiobjetivo (SisPG)

A Programação Genética (PG) multiobjetivo combinada com indicadores da Análise Técnica aplicada ao mercado de capitais desenvolvida neste trabalho se mostrou promissora. O SisPG, sistema desenvolvido neste trabalho, combinou otimização multiobjetivo, Programação Genética, Análise Técnica, Seleção de Características e Comitê de Decisão, visando identificar momentos adequados para a execução das ordens de compra e de venda. Até onde o autor tem conhecimento, tal combinação é única na literatura. A Programação Genética com a Análise Técnica já foi anteriormente tratada na literatura, porém, não tinha sido abordada uma implementação multiobjetivo.

O sistema foi testado em seis ações da BOVESPA (BBAS3, BOVA11, CMIG4, EMBR3, GGBR4 e VALE5), por dois períodos: (i) fevereiro de 2013 a fevereiro de 2015; (ii) julho de 2015 até julho de 2016. Os resultados alcançados foram muito promissores. O sistema obteve retornos financeiros consideravelmente acima do preço de variação das ações no mesmo período, superando as outras duas estratégias de investimento automatizado. Além disso, o sistema foi capaz de obter lucros significativos, mesmo em situações de forte depreciação do ativo, o que é um resultado notável.

6.1.3 Adaptação a outros mercados de capitais

Nos dois sistemas propostos, os métodos empregados (Análise Técnica, filtragem de *Outliers*, Seleção de Características/Seleção de Regras, Programação Genética e Comitê de Decisão) não se baseiam na natureza do mercado de ações e podem ser facilmente aplicados a outros tipos de séries temporais financeiras. O autor acredita que as únicas mudanças necessárias seriam nos parâmetros do método, da tabela 33. O ajuste dos parâmetros *StopGain* e *StopLoss*

devem receber atenção especial, levando em conta que os valores ajustados são adequados para o mercado brasileiro, que apresenta alta volatilidade.

6.1.4 Módulo de Negociação

Na maioria das publicações revisadas, observou-se que é dada pouca atenção ao módulo de negociação. Os artigos concentram-se principalmente na previsão das séries temporais. Neste trabalho, foi detalhada a forma de efetuar a compra e a venda de ações, não tendo sido encontrado na literatura sistema de negociação semelhante.

6.1.5 Regras Lógicas e Indicadores da Análise Técnica

Nos dois sistemas foram criados um conjunto de regras lógicas a partir de indicadores da AT. Essa decisão facilitou significativamente a construção das operações de mutação e cruzamento realizadas com a PG, uma vez que não foi necessário lidar com infactibilidades.

6.1.6 Estudos sobre a Bovespa

Outra contribuição importante são os testes destas implementações com ações da BOVESPA. Nos trabalhos de (ATSALAKIS; VALAVANIS, 2009a), (CAVALCANTE et al., 2016), e de (RATHER; SASTRY; AGARWAL, 2017) poucos métodos utilizam a BOVESPA como teste.

6.2 Proposta de Continuidade

Para dar continuidade a este trabalho, segue uma lista de sugestões:

- o sistema foi experimentado utilizando *candles* diários em operações de *SwingTrade*. Poderiam ser feitos experimentos com *candles* de 5, 10 e 15 minutos e testadas operações de *DayTrade*;
- ampliar o número de ações testadas e ajustar o sistema para outros mercados financeiros;
- ampliar o número de testes com os indicadores da AT;
- propor um método automatizado para ajustar os parâmetros de *StopGain* e *StopLoss* para se adaptar a diferentes mercados.

6.3 Produção Bibliográfica Durante o Doutorado

Segue a produção bibliográfica:

Periódicos

- A. Pimenta, E. G. Carrano, F. Gadelha Guimaraes, and L. Nametala, C. Aparecido. An Automated Investing Method for Stock Market Based on Multiobjective Genetic Programming. Computational Economics. Springer. **Artigo Publicado.**

Conferências

- C. A. Nametala, A. Pimenta, A. Pereira, and E. G. Carrano. An automated investment strategy using artificial neural networks and econometric predictors. In Proceedings of the XII Brazilian Symposium on Information Systems on Brazilian Symposium on Information Systems: Information Systems in the Cloud Computing Era-Volume 1, page 21. Brazilian Computer Society, 2016.
- A. Pimenta, E. G. Carrano, F. Gadelha Guimaraes, L. Nametala, C. Aparecido, and R. H. Takahashi. Goldminer: A genetic programming based algorithm applied to brazilian stock market. In Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on, pages 397–402. IEEE, 2014.

Referências

- ABBASS, H. A. A memetic pareto evolutionary approach to artificial neural networks. *Lecture Notes in Artificial Intelligence*, v. 2256, p. 1–12, 2001. Citado na página 90.
- ABBASS, H. A. Pareto neuro-evolution: Constructing ensemble of neural networks using multi-objective optimization. In: IEEE. *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on*. [S.l.], 2003. v. 3, p. 2074–2080. Citado na página 49.
- ACHELIS, S. B. *Technical Analysis from A to Z*. [S.l.]: McGraw Hill New York, 2001. Citado 12 vezes nas páginas 7, 17, 18, 20, 21, 22, 23, 24, 25, 27, 28 e 57.
- ALDRIDGE, I. *High-frequency trading: a practical guide to algorithmic strategies and trading systems*. [S.l.]: John Wiley & Sons, 2013. Citado 2 vezes nas páginas 2 e 3.
- ALFARO-CID, E.; SHARMAN, K.; ESPARCIA-ALCÁZAR, A. I. Genetic programming and serial processing for time series classification. *Evolutionary computation*, MIT Press, v. 22, n. 2, p. 265–285, 2014. Citado na página 14.
- ALLEN, F.; KARJALAINEN, R. Using genetic algorithms to find technical trading rules. *Journal of financial Economics*, Elsevier, v. 51, n. 2, p. 245–271, 1999. Citado 3 vezes nas páginas 12, 14 e 15.
- ATSALAKIS, G. S. Surveying stock market forecasting techniques, part i: Conventional methods. *Computational optimization in economics and finance research compendium.*, 2009b. Citado 2 vezes nas páginas 7 e 11.
- ATSALAKIS, G. S.; VALAVANIS, K. P. Surveying stock market forecasting techniques - part ii: Soft computing methods. *Expert Systems with Applications*, v. 36, n. 3, p. 5932–5941, 2009a. ISSN 0957-4174. Times Cited: 53 54. Disponível em: <<GotoISI>://WOS:000263817100024>. Citado 5 vezes nas páginas 5, 6, 12, 13 e 114.
- AUGUSTO, D. A.; BARBOSA, H. J. Programação genética. In: _____. *Manual de computação evolutiva e metaheurística*. [S.l.]: Imprensa da Universidade de Coimbra/Coimbra University Press, 2013. p. 67–86. Citado 4 vezes nas páginas 37, 38, 39 e 40.
- BACHELIER, L. *Théorie de la spéculation*. [S.l.]: Gauthier-Villars, 1900. Citado na página 4.
- BARNETT, V.; LEWIS, T. Outliers in statistical data: Chichester. *John Wiley&Sons*, 1994. Citado na página 32.
- BARROS, R. C. et al. A survey of evolutionary algorithms for decision-tree induction. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, IEEE, v. 42, n. 3, p. 291–312, 2012. Citado na página 14.
- BAZERMAN, M. H.; MOORE, D. A. Judgment in managerial decision making. Wiley, 2008. Citado na página 5.
- BMFBOVESPA. *Regulamentos e manuais*. 2017. <http://www.bmfbovespa.com.br/pt_br/regulacao/regulamentos-e-manuais/>. 03 Jul. 2017. Citado na página 3.

- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996. Citado na página 48.
- BREIMAN, L. Pasting small votes for classification in large databases and on-line. *Machine Learning*, Springer, v. 36, n. 1, p. 85–103, 1999. Citado na página 48.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado na página 48.
- BURGES, C. J. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, Springer, v. 2, n. 2, p. 121–167, 1998. Citado na página 12.
- CAMPELO, F. et al. A novel multiobjective immune algorithm using nondominated sorting. In: *11th International IGTE symposium on numerical field calculation in electrical engineering, Seggau, Austria*. [S.l.: s.n.], 2004. p. 308–313. Citado na página 40.
- CARRANO, E. G.; WANNER, E. F.; TAKAHASHI, R. H. C. A multi-criteria statistical based comparison methodology for evaluating evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, v. 15, p. 848–870, 2011. Citado na página 104.
- CAVALCANTE, R. C. et al. Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, Elsevier, v. 55, p. 194–211, 2016. Citado 5 vezes nas páginas 5, 6, 12, 13 e 114.
- CHANDRA, A.; YAO, X. Evolving hybrid ensembles of learning machines for better generalisation. *Neurocomputing*, Elsevier, v. 69, n. 7, p. 686–700, 2006. Citado na página 49.
- CHEN, C. Neural networks for financial market prediction. In: *IEEE. Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*. [S.l.], 1994. v. 2, p. 1199–1202. Citado na página 12.
- CHIARA, M. M. et al. Estratégia de arbitragem entre ações brasileiras e suas adrs: a resposta dos dados intraday. *Revista Brasileira de Economia de Empresas/Brazilian Journal of Business Economics*, v. 12, n. 2, 2012. Citado na página 2.
- CLEVELAND, W. S. Lowess: A program for smoothing scatterplots by robust locally weighted regression. *American Statistician*, JSTOR, p. 54–54, 1981. Citado 2 vezes nas páginas 83 e 84.
- COELLO, C. A. C.; LAMONT, G. B.; VELDHUIZEN, D. A. V. *Evolutionary algorithms for solving multi-objective problems*. [S.l.]: Springer, New York, 2007. Citado na página 41.
- COOK, R. D.; HAWKINS, D. M. Unmasking multivariate outliers and leverage points: comment. *Journal of the American Statistical Association*, JSTOR, v. 85, n. 411, p. 640–644, 1990. Citado na página 83.
- CORMEN, T. H. et al. *Introduction to Algorithms*. 3rd ed.. ed. [S.l.]: The MIT Press, 2009. Citado na página 90.
- CORNE, D. W. et al. Pesa-ii: Region-based selection in evolutionary multiobjective optimization. In: *CITeseer. Proceedings of the genetic and evolutionary computation conference (GECCO'2001)*. [S.l.], 2001. Citado na página 40.
- CORNE, D. W.; KNOWLES, J. D.; OATES, M. J. The pareto envelope-based selection algorithm for multiobjective optimization. In: *SPRINGER. Parallel Problem Solving from Nature PPSN VI*. [S.l.], 2000. p. 839–848. Citado na página 40.

- CORTEZ, P. *Modelos Inspirados na Natureza para a Previsão de Séries Temporais*. 2002. 188 f. Tese (Doutorado) — Tese (Doutorado em Informática)—Departamento de Informática, Universidade do Minho, Braga, 2002. Citado na página 14.
- DABHI, V. K.; CHAUDHARY, S. Financial time series modeling and prediction using postfix-gp. *Computational Economics*, Springer, p. 1–35, 2015. Citado 2 vezes nas páginas 14 e 15.
- DASARATHY, B. V.; SHEELA, B. V. A composite classifier system design: concepts and methodology. *Proceedings of the IEEE*, IEEE, v. 67, n. 5, p. 708–713, 1979. Citado na página 47.
- DEB, K. *Multi-objective optimization using evolutionary algorithms*. [S.l.]: John Wiley & Sons, 2001. v. 16. Citado 3 vezes nas páginas 41, 42 e 43.
- DEB, K.; JAIN, H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, IEEE, v. 18, n. 4, p. 577–601, 2014. Citado na página 41.
- DEB, K.; MIETTINEN, K. *Multiobjective optimization: interactive and evolutionary approaches*. [S.l.]: Springer Science & Business Media, 2008. v. 5252. Citado na página 41.
- DEB, K. et al. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, IEEE, v. 6, n. 2, p. 182–197, 2002. Citado 9 vezes nas páginas 41, 43, 44, 45, 56, 60, 70, 83 e 90.
- DONG, M.; ZHOU, X.-S. Exploring the fuzzy nature of technical patterns of us stock market. *Proceedings of Fuzzy System and Knowledge Discovery*, v. 1, p. 324–328, 2002. Citado na página 12.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012. Citado na página 34.
- ELDER, A. *Trading for a living: psychology, trading tactics, money management*. [S.l.]: John Wiley & Sons, 1993. v. 31. Citado 20 vezes nas páginas 3, 4, 7, 10, 17, 18, 20, 21, 22, 23, 24, 25, 27, 28, 29, 31, 55, 57, 63 e 82.
- ELDER, A. Como se transformar em um operador e investidor de sucesso. *Rio de Janeiro: Campus*, 2004. Citado na página 5.
- ESPEJO, P. G.; VENTURA, S.; HERRERA, F. A survey on the application of genetic programming to classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, v. 40, n. 2, p. 121–144, 2010. Citado na página 14.
- FABOZZI, F.; FOCARDI, S. M.; JONAS, C. High-frequency trading: Methodologies and market impact. *Review of Futures Markets*, v. 9, p. 7–38, 2011. Citado na página 2.
- FAMA, E. F. Efficient capital markets: A review of theory and empirical work*. *The journal of Finance*, Wiley Online Library, v. 25, n. 2, p. 383–417, 1970. Citado 2 vezes nas páginas 4 e 5.
- FONSECA, C. M.; FLEMING, P. J. An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary computation*, MIT Press, v. 3, n. 1, p. 1–16, 1995. Citado na página 40.

- FOREX. *FOREX*. 2014. <<http://pt.wikipedia.org/wiki/Forex>>. 05 nov. 2014. Citado na página 15.
- FOX, A. J. Outliers in time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 350–363, 1972. Citado na página 32.
- FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. In: SPRINGER. *European conference on computational learning theory*. [S.l.], 1995. p. 23–37. Citado na página 48.
- GRUBBS, F. E. Procedures for detecting outlying observations in samples. *Technometrics*, Taylor & Francis, v. 11, n. 1, p. 1–21, 1969. Citado na página 32.
- GUMBEL, E. Discussion on rejection of outliers by anscombe. *FJ Technometrics*, v. 2, p. 165–166, 1960. Citado 2 vezes nas páginas 16 e 32.
- HANSEN, L. K.; SALAMON, P. Neural network ensembles. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, v. 12, n. 10, p. 993–1001, 1990. Citado na página 47.
- HAWKINS, D. M. *Identification of outliers*. [S.l.]: Springer, 1980. v. 11. Citado na página 32.
- HORN, J.; NAFPLIOTIS, N.; GOLDBERG, D. E. A niched pareto genetic algorithm for multiobjective optimization. In: IEEE. *Evolutionary Computation, 1994. IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on*. [S.l.], 1994. p. 82–87. Citado na página 40.
- HOWARD, J. *Darwin*. [S.l.]: Edicoes Loyola, 1982. Citado na página 36.
- HUBERT, M.; VANDERVIEREN, E. An adjusted boxplot for skewed distributions. *Computational statistics & data analysis*, Elsevier, v. 52, n. 12, p. 5186–5201, 2008. Citado na página 33.
- JACOBS, R. et al. ^aadaptive mixtures of local experts, ^o neural computation, vol. 3. 1991. Citado 2 vezes nas páginas 48 e 49.
- KATTAN, A.; FATIMA, S.; ARIF, M. Time-series event-based prediction: An unsupervised learning framework based on genetic programming. *Information Sciences*, Elsevier, v. 301, p. 99–123, 2015. Citado na página 14.
- KIMOTO, T. et al. Stock market prediction system with modular neural networks. In: IEEE. *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*. [S.l.], 1990. p. 1–6. Citado na página 12.
- KIRILENKO, A. A.; LO, A. W. Moore’s law versus murphy’s law: Algorithmic trading and its discontents. *The Journal of Economic Perspectives*, American Economic Association, v. 27, n. 2, p. 51–72, 2013. Citado na página 1.
- KNOWLES, J.; CORNE, D. The pareto archived evolution strategy: A new baseline algorithm for pareto multiobjective optimisation. In: IEEE. *Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on*. [S.l.], 1999. v. 1. Citado na página 40.
- KOZA, J. R. *Genetic programming: on the programming of computers by means of natural selection*. [S.l.]: MIT press, 1992. v. 1. Citado 7 vezes nas páginas 13, 36, 37, 38, 39, 43 e 90.

- LARSON, R.; FARBER, B.; PATARRA, C. tradução técnica. *Estatística aplicada*. [S.l.]: Prentice Hall, 2004. Citado na página 33.
- LINDEN, R. *Algoritmos Genéticos (2a edicao)*. [S.l.]: Brasport, 2008. Citado na página 38.
- LIU, Y.; YAO, X.; HIGUCHI, T. Evolutionary ensembles with negative correlation learning. *IEEE Transactions on Evolutionary Computation*, IEEE, v. 4, n. 4, p. 380–387, 2000. Citado na página 49.
- LUH, G.-C.; CHUEH, C.-H.; LIU, W.-W. Moia: multi-objective immune algorithm. *Engineering Optimization*, Taylor & Francis, v. 35, n. 2, p. 143–164, 2003. Citado na página 40.
- LUKE, S. et al. A java-based evolutionary computation research system. *Online (March 2004)* <http://cs.gmu.edu/~eclab/projects/ecj>, 2004. Citado na página 15.
- MAJHI, R. et al. Prediction of s&p 500 and djia stock indices using particle swarm optimization technique. In: IEEE. *Evolutionary Computation, 2008. CEC 2008.(IEEE World Congress on Computational Intelligence). IEEE Congress on*. [S.l.], 2008. p. 1276–1282. Citado na página 12.
- MARKOWITZ, H. Portfolio selection. *The journal of finance*, Wiley Online Library, v. 7, n. 1, p. 77–91, 1952. Citado na página 1.
- MARQUES, F. C.; GOMES, R. M. Análise de séries temporais aplicadas ao mercado financeiro com o uso de algoritmos genéticos e lógica nebulosa. *VII Encontro Nacional de Inteligência Artificial. Bento Gonçalves: Sociedade Brasileira de Computação*, p. 749–768, 2009. Citado na página 12.
- MORETTIN, P. A. *Econometria Financeira*. [S.l.]: Blucher, 2011. Citado na página 11.
- MURPHY, J. J. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. [S.l.]: Penguin, 1999. Citado 20 vezes nas páginas 3, 4, 7, 8, 9, 11, 17, 18, 20, 21, 22, 23, 24, 25, 27, 28, 29, 31, 57 e 63.
- MYSZKOWSKI, P. B.; RACHWALSKI, Ł. Trading rule discovery on warsaw stock exchange using coevolutionary algorithms. In: *Proceedings of the International Multiconference on Computer Science and Information Technology*. [S.l.: s.n.], 2009. v. 3, p. 81–88. Citado 3 vezes nas páginas 14, 15 e 16.
- NORONHA, M. Análise técnica: teorias, ferramentas, estratégias. *São Paulo: Editec*, 1995. Citado na página 9.
- OLTEAN, M. et al. Genetic programming with linear representation: a survey. *International Journal on Artificial Intelligence Tools*, World Scientific, v. 18, n. 02, p. 197–238, 2009. Citado na página 36.
- OPITZ, D. W.; SHAVLIK, J. W. Actively searching for an effective neural network ensemble. *Connection Science*, Taylor & Francis, v. 8, n. 3-4, p. 337–354, 1996. Citado na página 47.
- PERRONE, M. P.; COOPER, L. N. *When networks disagree: Ensemble methods for hybrid neural networks*. [S.l.], 1992. Citado na página 47.
- PIMENTA, A. et al. Goldminer: A genetic programming based algorithm applied to brazilian stock market. In: IEEE. *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on*. [S.l.], 2014. p. 397–402. Citado 2 vezes nas páginas 14 e 15.

POLI, R. et al. *A field guide to genetic programming*. [S.l.]: Lulu. com, 2008. Citado na página 14.

POLIKAR, R. Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, IEEE, v. 6, n. 3, p. 21–45, 2006. Citado 2 vezes nas páginas 47 e 48.

POTVIN, J.-Y.; SORIANO, P.; VALLÉE, M. Generating trading rules on the stock markets with genetic programming. *Computers & Operations Research*, Elsevier, v. 31, n. 7, p. 1033–1047, 2004. Citado 3 vezes nas páginas 14, 15 e 16.

RATHER, A. M.; SASTRY, V.; AGARWAL, A. Stock market prediction and portfolio selection models: a survey. *OPSEARCH*, Springer, p. 1–22, 2017. Citado 5 vezes nas páginas 5, 6, 12, 13 e 114.

ROKACH, L. Ensemble-based classifiers. *Artificial Intelligence Review*, Springer, v. 33, n. 1, p. 1–39, 2010. Citado 2 vezes nas páginas 47 e 48.

RUSKIN, J. Ad valorem. *Unto This Last, and other writings*, p. 204–228, 1860. Citado na página 5.

SÁ, G. T. de. *Administração de investimentos: teoria de carteiras e gerenciamento do risco*. [S.l.]: Qualitymark, 1999. Citado na página 11.

SAMUELSON, P. A. Proof that properly anticipated prices fluctuate randomly. *IMR; Industrial Management Review (pre-1986)*, Massachusetts Institute of Technology, Cambridge, MA, v. 6, n. 2, p. 41, 1965. Citado na página 5.

SCHAFFER, J. D. Multiple objective optimization with vector evaluated genetic algorithms. In: L. ERLBAUM ASSOCIATES INC. *Proceedings of the 1st international Conference on Genetic Algorithms*. [S.l.], 1985. p. 93–100. Citado 2 vezes nas páginas 39 e 40.

SCHALLER, R. R. Moore's law: past, present and future. *IEEE spectrum*, IEEE, v. 34, n. 6, p. 52–59, 1997. Citado na página 1.

SCHAPIRE, R. E. The strength of weak learnability. *Machine learning*, Springer, v. 5, n. 2, p. 197–227, 1990. Citado 2 vezes nas páginas 47 e 48.

SHLEIFER, A. *Inefficient markets: An introduction to behavioural finance*. [S.l.]: OUP Oxford, 2000. Citado na página 5.

SILVA, S. F. d. *Seleção de características por meio de algoritmos genéticos para aprimoramento de rankings e de modelos de classificação*. Tese (Doutorado) — Universidade de São Paulo, 2011. Citado na página 35.

SRINIVAS, N.; DEB, K. Muultiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, MIT Press, v. 2, n. 3, p. 221–248, 1994. Citado na página 40.

TÉCNICASTRADING. *Técnicas de Trading*. 2017. <<https://http://www.tecnicasdetrading.com/2012/02/oscilador-de-chaikin.html>>. 14 jan. 2017. Citado na página 23.

TELLER, A. Evolving programmers: The co-evolution of intelligent recombination operators. In: MIT PRESS. *Advances in genetic programming*. [S.l.], 1996. p. 45–68. Citado na página 36.

THOMPSON, K. J.; SIGMAN, R. S. Statistical methods for developing ratio edit tolerances for economic censuses. In: *American Statistical Association, Proceedings of the Section on Survey Research Methods*. [S.l.: s.n.], 1996. p. 166–171. Citado na página 33.

TUKEY, J. W. *Exploratory data analysis*. Reading, Mass., 1977. Citado 2 vezes nas páginas 32 e 33.

VASILAKIS, G. A. et al. A genetic programming approach for eur/usd exchange rate forecasting and trading. *Computational Economics*, Springer, v. 42, n. 4, p. 415–431, 2013. Citado na página 14.

WEBER, M. *The Protestant ethic and the spirit of capitalism*. [S.l.]: Allen & Unwin, 1930. Citado na página 5.

WIKIPEDIA. *Processo de impeachment de Dilma Rousseff*. 2016. <http://pt.wikipedia.org/wiki/Processo_de_impeachment_de_Dilma_Rousseff/>. 26 nov. 2016. Citado na página 102.

WIKIPEDIA. *Ibovespa*. 2017. <http://pt.wikipedia.org/wiki/Bolsa_de_Valores,_Mercadorias_e_Futuros_de_S~ao_Paulo>. 03 Jun. 2017. Citado na página 92.

WIKIPEDIA. *Lei da oferta e da procura*. 2017. <http://pt.wikipedia.org/wiki/Lei_da_oferta_e_da_procura/>. 03 junho. 2016. Citado na página 2.

XP. *XP Investimentos*. 2017. <<https://www.xpi.com.br/>>. 13 jan. 2017. Citado 14 vezes nas páginas 11, 12, 10, 20, 21, 22, 24, 25, 26, 27, 29, 30, 31 e 32.

YOSHINAGA, C. E. et al. Finanças comportamentais: uma introdução. *REGE Revista de Gestão*, v. 15, n. 3, p. 25–35, 2008. Citado na página 5.

ZHOU, A. et al. Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*, Elsevier, v. 1, n. 1, p. 32–49, 2011. Citado na página 39.

ZITZLER, E.; KÜNZLI, S. Indicator-based selection in multiobjective search. In: SPRINGER. *International Conference on Parallel Problem Solving from Nature*. [S.l.], 2004. p. 832–842. Citado na página 40.

ZITZLER, E. et al. *SPEA2: Improving the strength Pareto evolutionary algorithm*. [S.l.]: Eidgenössische Technische Hochschule Zürich (ETH), Institut für Technische Informatik und Kommunikationsnetze (TIK), 2001. Citado na página 40.

ZITZLER, E. et al. *An evolutionary algorithm for multiobjective optimization: The strength pareto approach*. [S.l.]: Citeseer, 1998. v. 43. Citado na página 40.