

**UNIVERSIDADE FEDERAL DE ALFENAS**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

*Carlos Augusto N. De C. Rivilino*  
*Thiago de Oliveira Almeida*

**MAXIMUS II: UM DESCOBRIDOR DE REGRAS**  
**TÉCNICAS PARA O MERCADO ACIONÁRIO DA**  
**BOLSA DE VALORES DE SÃO PAULO**

Alfenas, 15 de Junho de 2011.



**UNIVERSIDADE FEDERAL DE ALFENAS**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**MAXIMUS II: UM DESCOBRIDOR DE REGRAS**  
**TÉCNICAS PARA O MERCADO ACIONÁRIO DA**  
**BOLSA DE VALORES DE SÃO PAULO**

*Carlos Augusto N. De C. Rivilino*

*Thiago de Oliveira Almeida*

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal de Alfenas como requisito parcial para obtenção do Título de Bacharel em Ciência da Computação.

Orientador: Prof. Humberto César Brandão de Oliveira

Alfenas, 15 de Junho de 2011.

*Carlos Augusto N. De C. Rivilino*

*Thiago de Oliveira Almeida*

**MAXIMUS II: UM DESCOBRIDOR DE REGRAS  
TÉCNICAS PARA O MERCADO ACIONÁRIO DA  
BOLSA DE VALORES DE SÃO PAULO**

A Banca examinadora abaixo-assinada aprova a monografia apresentada como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação pela Universidade Federal de Alfenas.

---

**Prof. Humberto César Brandão de Oliveira (Orientador)**

**Universidade Federal de Alfenas**

---

**Prof. Luiz Eduardo da Silva**

**Universidade Federal de Alfenas**

---

**Prof. Nelson José Freitas da Silveira**

**Universidade Federal de Alfenas**

Alfenas, 15 de Junho de 2011.

Dedicamos este trabalho a todos os docentes do curso da Computação da UNIFAL, cujos trabalhos foram merecidamente reconhecidos em 2011 pelo Ministério da Educação.



# AGRADECIMENTO

Eu, Carlos Augusto, agradeço aos meus pais, Natanael Rivilino e Grace Amaral Nantes de Castilho, por todo apoio, confiança e amor que me deram durante minha vida.

Agradeço a minha avó, Teresinha Amaral, por possibilitar meu prosseguimento neste curso.

Agradeço a minha prima, Flávia Campos Nantes, por ter me motivado a graduar na UNIFAL.

Agradeço ao corpo docente pelos ensinamentos e pelo esforço para tornar o curso da computação melhor.

Agradeço aos integrantes do Laboratório de Pesquisa e Desenvolvimento (LP&D), por tornarem minhas tardes no laboratório animadoras e divertidas.

Eu, Thiago de Oliveira, agradeço em primeiro lugar aos meus pais por me suportar em todas as minhas decisões. Esse apoio foi vital para todas as conquistas da minha vida.

Depois agradeço a todos os docentes que me proveram com um curso de excelência, estou muito honrado por essa oportunidade.

Por fim, agradeço a todos aqueles que cruzaram o meu caminho durante essa trajetória.

Agradecemos ao professor, orientador e amigo Humberto por nos dar suporte em Belo Horizonte para aprendizado na corretora Corval, pelos ensinamentos e por viabilizar um excelente local de trabalho e pesquisa.

"The best theory is inspired by practice. The best practice is inspired by theory."

Donald E. Knuth (1975)



## RESUMO

No mercado, é comum a análise dos dados históricos de ações e o relacionamento entre os diversos indicadores de mercado para a tomada de decisão. Tais decisões compreendem quais ações comprar, em que momento comprá-las e por quanto tempo mantê-las a fim de aumentar seu capital em carteira. Atualmente a Internet provê fontes com grande quantidade destes dados, facilitando o uso de técnicas de aprendizado de máquina para resolver estes problemas de decisão. Este trabalho investiga a possibilidade de se encontrar regras técnicas lucrativas no mercado acionário da BOVESPA, utilizando a estrutura de árvores de decisões binárias, técnica de Programação Genética e o método *Hill-climbing*. A Programação Genética e o método *Hill-climbing* foram utilizados para encontrar agentes financeiros artificiais lucrativos para as ações PETR4.SA, BBDC.SA e VALE5.SA. As árvores de decisões binárias foram utilizadas como estrutura de dados para representar as regras técnicas dos agentes artificiais encontrados e para facilitar a visualização das mesmas pelo usuário.

**Palavras-Chave:** mercado acionário, investimento, programação genética, regras técnicas.

## **ABSTRACT**

In the stock market some investors analyze the historical data of stocks and the relationship between the various market indicators to make decisions. These decisions include what stock to buy, when to buy them and how long to keep them in order to increase their capital in their portfolio. Currently the Internet provides sources with large amounts of data, facilitating the use of machine learning techniques to solve these decision problems. This work investigates the possibility of finding technical rules of the BOVESPA stock market, using the structure of binary decision trees, techniques of genetic programming and hill-climbing method. The Genetic Programming and Hill-climbing method was used to find profitable artificial financial agents for the stocks PETR4.SA, BBDC.SA and VALE5.SA. Binary decision trees were used as data structure to represent the technical rules of artificial agents and for facilitating their viewing for user.

**Keyword:** stock market, investment, genetic programming, technical regulations.

## LISTA DE FIGURAS

Figura 1 - Representação de um candle.....	28
Figura 2 - Gráfico <i>candlestick</i> do preço da ação PETR4.SA. Cada <i>candle</i> tem período de 1 dia. ....	28
Figura 3 - A figura mostra quatro <i>drawdowns</i> baseado em Sornette(2004), pág 52 .	29
Figura 4 - Árvore de decisão binária.....	32
Figura 5 - Representação de uma possível árvore de compra ou de venda.....	45
Figura 6 - Gráfico do preço do período de teste número 1.....	56
Figura 7 - Gráfico de desempenho do período 1.....	57
Figura 8 - Gráfico do preço do período de teste número 2.....	58
Figura 9 - Gráfico de desempenho do período 2.....	59
Figura 10 - Gráfico do preço do período de teste número 3. ....	60
Figura 11 - Gráfico de desempenho do período 3.....	61
Figura 12 - Gráfico do preço do período de teste número 1. ....	62
Figura 13 - Gráfico de desempenho do período 1.....	63
Figura 14 - Gráfico do preço do período de teste número 2.....	64
Figura 15 - Gráfico de desempenho do período 2.....	65
Figura 16 - Gráfico do preço do período de teste número 3.....	66
Figura 17 - Gráfico de desempenho do período 3.....	67

## LISTA DE TABELAS

Tabela 1 - Custos de corretagem segundo a tabela BOVESPA.....	26
Tabela 2 - Emolumento cobrado de operações no mercado acionário .....	26
Tabela 3 - Custos de Manutenção de Contas de Custódia. ....	27
Tabela 4 - Estrutura do arquivo com dados da PETR4. ....	42
Tabela 5 - Período total de cada base de dados .....	42
Tabela 6 - Processo de redução de candles. Reduz-se 10 <i>candles</i> de 1 minuto para 1 <i>candle</i> de 10 minutos, no arquivo da PETR4. A última linha da tabela é o resultado de um processamento parcial. ....	44
Tabela 7 - Conjunto de regras da gramática G .....	46
Tabela 8 - Conjunto de funções usadas em <EXP_A>. ....	48
Tabela 9 - Parâmetros para o algoritmo genético.....	51
Tabela 10 - Parâmetros para função objetivo .....	52
Tabela 11 - Parâmetros para agente financeiro .....	53
Tabela 12 - Parâmetros para <i>hill-climbing</i> .....	53
Tabela 13 - Descrição do período de teste número 1. ....	56
Tabela 14 - Desempenho dos agentes do período 1. ....	57
Tabela 15 - Descrição do período de teste número 2. ....	58
Tabela 16 - Desempenho dos agentes do período 2. ....	59
Tabela 17 - Descrição do período de teste número 3. ....	60
Tabela 18 - Desempenho dos agentes do período 3. ....	61
Tabela 19 - Descrição do período de teste número 1. ....	62
Tabela 20 - Desempenho dos agentes do período 1. ....	63
Tabela 21 - Descrição do período de teste número 2. ....	64
Tabela 22 - Desempenho dos agentes do período 2. ....	65
Tabela 23 - Descrição do período de teste número 3. ....	66
Tabela 24 - Desempenho dos agentes do período 3. ....	67
Tabela 25 - Desempenho geral dos agentes .....	68

## LISTA DE ALGORITMOS

Algoritmo 1 - Esquema geral de um algoritmo evolucionário (EIBEN, 2003) .....	37
Algoritmo 2 - Algoritmo de <i>hill-climbing</i> extraído de Russel e Norvig (2004). .....	39

# LISTA DE ABREVIACOES

BOVESPA	Bolsa de Valores de So Paulo
CBLC	Companhia Brasileira de Liquidao e Custdia
CONST	Constante
IFR	ndice de Fora Relativa
LBB	Linha Inferior do <i>Bollinger Bands</i>
MACD	Mdia Mvel Convergncia/Divergncia
MM	Mdia Mvel
MME	Mdia Mvel Exponencial
MMS	Mdia Mvel Simples
OSCE	Oscilador Estocstico
PABERT	Preo de Abertura
PFECH	Preo de Fechamento
PMAX	Preo Mximo
PMIN	Preo Mnimo
PROC	<i>Price Rate Of Change</i>
UBB	Linha Superior do <i>Bollinger Bands</i>
VABERT	Volume de Abertura
VFECH	Volume de Fechamento
VMAX	Volume Mximo
VMIN	Volume Mnimo
WILR	Williams %R

# SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>17</b>
1.1 JUSTIFICATIVA E MOTIVAÇÃO .....	18
1.2 PROBLEMATIZAÇÃO .....	19
1.3 OBJETIVOS .....	19
1.3.1 Geral .....	19
1.3.2 Específicos .....	19
1.4 ORGANIZAÇÃO DA MONOGRAFIA .....	20
<b>2 MERCADO ACIONÁRIO</b> .....	<b>21</b>
2.1 INTRODUÇÃO .....	21
2.2 MODELOS DE ANÁLISE DE MERCADO .....	22
2.2.1 Modelo de Análise Fundamentalista .....	22
2.2.2 Modelo de Análise Técnica .....	23
2.3 CUSTOS OPERACIONAIS .....	25
2.4 OUTROS CONCEITOS DE MERCADO .....	27
2.4.1 <i>Candle</i> .....	27
2.4.2 <i>Drawdown</i> .....	29
<b>3 ALGORITMOS DE OTIMIZAÇÃO</b> .....	<b>31</b>
3.1 ÁRVORE DE DECISÃO .....	31
3.2 PROGRAMAÇÃO GENÉTICA .....	33
3.2.1 Indivíduo .....	34
3.2.2 População .....	35
3.2.3 Recombinação genética .....	36
3.2.3.1 <i>Crossover</i> .....	36
3.2.3.2 Reprodução .....	36
3.2.4 Mutação .....	37
3.2.5 Função objetivo .....	34
3.3 HILL-CLIMBING .....	38
<b>4 PROJETO</b> .....	<b>41</b>
4.1 METODOLOGIA .....	41
4.2 DEFINIÇÃO DA HIPÓTESE .....	41
4.3 DEFINIÇÃO E OBTENÇÃO DAS BASES DE DADOS .....	42
4.4 PROCESSAMENTO DAS BASES DE DADOS .....	43
4.5 DEFINIÇÃO DAS ESTRUTURAS DO PROGRAMA .....	45
4.5.1 Indivíduo .....	45
4.5.2 Função de avaliação .....	49
4.5.3 Operações de variação .....	50
4.6 DEFINIÇÃO DO ALGORITMO .....	51
4.6.1 Parâmetros do algoritmo .....	52
<b>5 RESULTADOS</b> .....	<b>55</b>
5.1 CONSIDERAÇÕES INICIAIS .....	55
5.2 RESULTADOS .....	55

5.2.1 Ações PETR4.SA.....	56
5.2.1.1 Período de teste número 1.....	56
5.2.1.2 Período de teste número 2.....	58
5.2.1.3 Período de teste número 3.....	60
5.2.2 Ações VALE.SA.....	62
5.2.2.1 Período de teste número 1.....	62
5.2.2.2 Período de testes número 2.....	64
5.2.2.3 Período de testes número 3.....	66
5.3 DESEMPENHO GERAL.....	68
<b>6 CONCLUSÃO.....</b>	<b>69</b>
6.1 COMPARAÇÕES.....	69
6.2 DISCUSSÕES E TRABALHOS FUTUROS.....	70
<b>7 REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>73</b>



# 1

## Introdução

*Neste capítulo serão apresentados ao leitor uma visão geral sobre o tema, a motivação e justificativa do projeto (Seção 1.1 Justificativa e Motivação), o domínio do problema a ser trabalhado (Seção 1.2 Problematização) e a lista de objetivos que se pretende atingir (Seção 1.3 Objetivos). A Seção 1.4 mostra como a monografia está organizada.*

Uma das possíveis opções de investimento de renda variável é o mercado acionário, no qual investidores tentam aumentar seu capital adquirindo ações a um baixo preço e vendendo-as quando este se eleva. Entretanto, a natureza instável e complexa do mercado torna essa tarefa difícil. Para auxiliar as decisões de investimento neste mercado de capitais é comum o uso de ferramentas que traçam gráficos e realizam o cálculo de indicadores. Além destas, algumas corretoras e investidores possuem ferramentas especializadas de apoio a decisão, como planilha de regras técnicas e *software* robô negociador que os indicam o momento correto de comprar ou vender ações de acordo com as estratégias de investimento inseridas nestas. Contudo, estas ferramentas exigem que o usuário inclua uma estratégia de investimento apropriada para se obter lucro.

Independente do perfil do investidor, uma estratégia de investimento apropriada é aquela que resulta em lucro. Deve-se considerar ainda a quantidade de vezes que se negocia no mercado, devido à existência de taxas aplicadas a cada operação de compra ou venda realizada. Por exemplo, o número de operações realizadas por um investidor de curto prazo (*trader*) em um dado período de tempo pode ser maior que o número de operações efetuadas por um investidor de longo prazo no mesmo período. Portanto, é necessário que a quantidade média de operações efetuadas por uma estratégia de negociação seja parametrizada, cabendo ao investidor ajustá-la de acordo com seu perfil.

A procura de estratégias lucrativas no passado envolve a suposição de que quando empregadas em tempos atuais estas também obterão lucro. De fato, o modelo de análise

técnica de mercado supõe que padrões se repetem ao longo do tempo na curva de preço da ação. Este modelo incentiva a observação de dados históricos para o auxílio na tomada de decisão dos investidores.

Trabalhos precedentes (SCHULENBURG & ROSS, 1999, 2000, 2001) afirmam ser possível obter lucro no mercado acionário estrangeiro evoluindo estratégias de investimento com grupos de agentes financeiros artificiais. Estes operam na base de valores históricos de uma determinada ação. Embora, a inclusão da estratégia inicial ainda seja uma exigência nestes trabalhos, é possível observar que a máquina consegue criar outras estratégias apropriadas derivadas da primeira, o que reforça a pesquisa de ferramentas mais elaboradas.

Neste trabalho foram testadas e aplicadas as técnicas de Programação Genética, o método *Hill-Climbing* (subida na encosta) em conjunto com árvores de decisões binárias para a geração de regras técnicas lucrativas no mercado acionário da BOVESPA.

## 1.1 Justificativa e Motivação

O mercado acionário é um sistema complexo de ser modelado. Os investidores que atuam neste segmento utilizam diferentes modelos de análise para fundamentar suas decisões de compra e venda de ações. O modelo de análise técnica se apóia no estudo de gráficos e na relação entre os indicadores financeiros disponíveis. Existem diversas ferramentas (YAHOO FINANCE, TERMINAL ENFOQUE, GUIAINVEST) capazes de construir estes gráficos, restando ao investidor decidir como relacionar os indicadores calculados. Esta é uma tarefa complexa, já que uma grande quantidade dos indicadores possui parâmetros variáveis gerando uma infinidade de estratégias possíveis de serem adotadas.

Este projeto se mostra importante para os investidores que adotam o modelo de análise técnica e para as corretoras. Para o primeiro, porque seria interessante o uso de um *software* descobridor de estratégias lucrativas. Para o segundo grupo, esta seria mais uma ferramenta para oferecer aos seus clientes e para uso em seus próprios investimentos.

## 1.2 Problematização

A complexidade do comportamento do mercado de ações traz dificuldades aos investidores na hora de tomar decisões. Variáveis como preço e volume de negociação e indicadores derivados destas variáveis podem ser relacionados para fundamentá-las. No entanto, avaliar estas relações sob pressão emocional e de forma eficiente é um desafio ao investidor. O presente projeto objetiva auxiliar este grupo social em seus negócios investigando as respostas para as seguintes questões:

- Quais possíveis relações entre os indicadores resultam em regras técnicas lucrativas?
- Os investimentos com as estratégias geradas pelo sistema superam o investimento de renda fixa e a estratégia *buy-and-hold* (compra-e-espera) no mesmo período?
- É viável a utilização de um robô inteligente para negociar no mercado acionário da BOVESPA, considerando os custos operacionais deste mercado?

## 1.3 Objetivos

### 1.3.1 Geral

O objetivo geral compreende o desenvolvimento de um *software* para geração de regras técnicas lucrativas no mercado acionário da BOVESPA, que descreva de forma compreensível as boas estratégias de investimento encontradas.

### 1.3.2 Específicos

Os objetivos específicos compreendem:

- Realizar um levantamento dos indicadores mais utilizados no modelo de análise técnica de mercado;
- Aplicar e testar diferentes modelos de aprendizado de máquina no mercado acionário da BOVESPA;
- Criar um conjunto de boas estratégias encontradas pelo *software* proposto;
- Fornecer um comparativo entre a lucratividade das estratégias empregadas pelos agentes financeiros artificiais gerados pelo programa e as estratégias de investimento comumente utilizadas pelos investidores – a saber, compra-e-espera (*buy-and-hold*) e investimento em banco.

## 1.4 Organização da Monografia

O restante deste trabalho está organizado na seguinte seqüência: o Capítulo 2 introduz os conceitos do mercado acionário necessários para o entendimento do problema trabalhado. O Capítulo 3 apresenta um referencial teórico sobre a estrutura e os algoritmos utilizados na busca de soluções para o problema. O Capítulo 4 mostra a metodologia seguida durante o desenvolvimento e as definições usadas no trabalho. O Capítulo 5 apresenta os gráficos comparativos do valor de carteira dos agentes encontrados pelo programa e dos agentes que utilizam as estratégias comumente empregadas pelos investidores. Finalmente, o Capítulo 6 mostra a conclusão, as discussões e as propostas para trabalhos futuros na área.

# 2

## Mercado acionário

*Este capítulo apresenta uma visão geral sobre o mercado acionário (Seção 2.1 Introdução), descreve os modelos de análise frequentemente utilizados na hora do investimento (Seção 2.2 Modelos de Análise de Mercado), detalha os custos operacionais envolvidos neste mercado (Seção 2.3 Custos Operacionais) e introduz os conceitos de mercado utilizados no restante deste trabalho (Seção 2.4 Outros conceitos de mercado).*

### 2.1 Introdução

As empresas, à medida que se expandem, carecem de recursos. Uma alternativa para obterem recursos financeiros é emitir títulos, chamados de ações, para negociação em bolsa de valores. “Ações são títulos de renda variável, emitidos por sociedades anônimas, que representam a menor fração do capital da empresa emitente” (BMF&BOVESPA, 2010). Comprando ações, o investidor se torna um co-proprietário da sociedade anônima da qual é acionista e participa dos seus resultados (BMF&BOVESPA, 2010). Isto fornece à empresa o capital necessário para seu crescimento e ao acionista a possibilidade de lucros se o desempenho da empresa no mercado for favorável e o valor das ações subirem.

Quando os investidores têm bons prospectos de uma companhia, fazem ofertas por suas ações, elevando os preços destas. Em contrapartida, no caso de um prospecto ruim vendem suas ações diminuindo o preço (ELDER, 2002). Isto é a lei da oferta e demanda, quanto maior a quantidade de compradores de uma ação, mais caro um vendedor desta ação poderá vendê-la, provocando uma tendência de aumento de preço da mesma. Por outro lado, quanto maior a quantidade de vendedores desta ação, ou seja, muita oferta, menor tende a ser o preço de venda. Contudo, escolher em qual empresa investir e por quanto tempo ser proprietário das ações, não é uma tarefa fácil. Na literatura econômica,

dois modelos de análise de mercado auxiliam os investidores nesta tomada de decisão. Ambos serão descritos na Seção 2.2.

Na Seção 2.3 serão detalhados os custos gerados aos investidores que negociam no mercado acionário, incluindo valores de custódia de ações, custos operacionais e impostos. Finalmente, na Seção 2.4 detalhamos os conceitos de mercado considerados nos métodos de otimização aplicados neste trabalho.

## **2.2 Modelos de Análise de Mercado**

Evidentemente, realizar compras e vendas no mercado acionário sem alguma estratégia bem definida e seguida com disciplina pode resultar em altas perdas para o investidor, com baixa probabilidade de recuperação do capital perdido. Por isso, um investidor bem instruído tende a seguir um dos modelos de análises de mercado, detalhados na literatura econômica: o modelo de análise fundamentalista e o modelo de análise técnica, introduzidos nas seções que seguem.

### **2.2.1 Modelo de Análise Fundamentalista**

O modelo de análise fundamentalista sugere que as ações sejam analisadas de acordo com a saúde financeira da companhia. Pode-se avaliar se a empresa apresentará um futuro promissor analisando seus fundamentos, como a relação preço/lucro, o lucro líquido, o grau de endividamento, o patrimônio líquido, o pagamento de dividendos, entre outros. A análise fundamentalista sofre grande influência pelas notícias divulgadas na mídia (PIAZZA, 2009).

Na prática, o investidor que segue este modelo de análise não comprará ações que estão supervalorizadas se comparado aos fundamentos de sua empresa emitente e procurará descobrir a ação que mais irá se valorizar até o fim de um ano. Além disso, o

investidor fundamentalista tende a aplicar seu capital e deixá-lo rendendo em um longo período de tempo, se mostrando interessado na empresa e acreditando no progresso da mesma em longo prazo.

### 2.2.2 Modelo de Análise Técnica

O modelo de análise técnica prevê o uso de dados históricos da ação para geração de gráficos e procura por padrões no preço, tentando reconhecer tendências de alta ou baixa em seus estágios iniciais, para identificar possíveis oportunidades de compra ou venda.

Segundo Achelis (2010), a análise técnica é baseada na análise de preço e volume e os campos que os definem são os seguintes:

- **Abertura:** este é o preço da primeira negociação do período. Quando se trata de dados diários o preço de abertura é o preço da primeira negociação realizada no dia. Se os dados estiverem agrupados de minuto em minuto, o preço de abertura será o preço da primeira negociação de um minuto.
- **Máximo:** é o maior valor que o preço da ação atingiu durante o período. É o ponto onde houve mais vendedores do que compradores, isto é, sempre existem vendedores desejando vender a preços mais altos, mas o preço máximo representa o maior preço que os compradores desejaram pagar.
- **Mínimo:** este é o menor valor que o preço da ação atingiu durante o período. É o ponto onde houve mais compradores que vendedores. Novamente, sempre existem compradores desejando comprar a preços menores, mas o preço mínimo representa o menor preço que os vendedores desejaram aceitar.
- **Fechamento:** consiste no último preço que a ação foi negociada durante o período. Devido a sua disponibilidade, o preço de fechamento é preço mais usado para análises.
- **Volume:** representa o número de ações que foram negociadas durante o período.

- **Bid:** é o preço que um comprador está disposto a pagar pela ação no momento no período *intraday*. Entenda-se por período *intraday* o período em que o mercado está aberto para negociações durante um dia de negociação e dados *intraday* os dados coletados durante subperíodos deste período.
- **Ask:** é o preço que um vendedor deseja receber para negociar a ação no momento (no período *intraday*).

Na prática, além destes dados brutos o investidor utiliza indicadores de mercado para reforçar suas decisões. Um indicador é um cálculo matemático que pode ser aplicado ao campo de preço ou volume da ação negociada. Alguns indicadores bastante usados são:

- **Médias móveis (MM):** mostram o valor médio dos dados em determinado período. Uma MM de cinco dias mostra o preço médio dos últimos cinco dias, uma MM de 20 dias mostra o preço médio dos últimos 20 dias. Quando se conectam as médias móveis de cada dia, cria-se uma curva de média móvel no gráfico. O valor da MM depende de dois fatores: os valores cuja média está sendo calculada e a amplitude do intervalo temporal analisado. As médias móveis funcionam como zonas de suporte e resistência. Quando em ascensão tende a atuar como piso sob a curva preços e a quando em declínio tende a funcionar como teto sobre os preços (ELDER, 2002).
- **Índice de Força Relativa (IFR):** é um indicador que assume valores de 0 a 100, usado para identificar situações de saturação de compra ou venda de uma ação ou índice. Em outras palavras, é usado para identificar oportunidades de compra durante o declínio em curto prazo de um ativo ou oportunidades de venda durante sua ascensão. Valores altos indicam saturação de compra e oportunidades para vendas. Números baixos indicam saturação de vendas e oportunidades para compras (FONTANILLS; GENTILE, 2001). Acredita-se que o mercado está próximo a um topo quando o índice alcança um valor maior que 70, portanto, tendendo a uma redução de preço. Em contrapartida acredita-se que abaixo de 30, existe uma tendência de alta (PIAZZA, 2009).



- ***Bollinger Bands***: indicador criado por John Bollinger que resulta em duas curvas traçadas acima e abaixo de uma média móvel no gráfico em níveis de desvio padrão, portanto suas curvas são auto-ajustáveis: se alargam durante períodos instáveis e se contraem em períodos mais calmos.

Além destes indicadores descritos aqui, existem diversos outros, descritos em Achelis (2010), que podem ser utilizados, devendo o investidor selecionar o que mais atender sua necessidade e utilizá-los paralelamente confirmando suas decisões.

## 2.3 Custos Operacionais

Operar no mercado acionário envolve custos, incluindo custo mensal pago para o serviço de custódia de ações, custo por operação realizada, taxas e imposto. Portanto, nos cálculos do lucro líquido é preciso analisar minuciosamente todos estes custos para verificar se as operações realizadas estão rendendo lucros reais, o que justifica a necessidade de descrevê-los.

O primeiro custo que o investidor possui neste mercado, está relacionado à execução de ordens, de compra e venda, enviadas à BOVESPA. As corretoras cobram uma taxa a cada ordem executada pelo investidor, chamada de taxa de corretagem. Normalmente as corretoras cobram valores entre 7 e 20 reais, quando a operação é feita via *homebroker*, uma plataforma online para investimento. Por outro lado, se a negociação for feita via mesa de corretagem, a taxa de corretagem é aplicada de acordo com a tabela BOVESPA (veja Tabela 1), que consiste em um valor fixo mais um percentual do valor que está sendo negociado. Além da taxa de corretagem, o investidor também paga uma taxa para a BOVESPA (emolumento) sobre o valor negociado e pode variar de acordo com o tipo de operação realizada. Operações *day trade*, que é a compra e venda de um ativo no mesmo dia, tem custo reduzido de emolumento (veja Tabela 2).

**Tabela 1 – Custos de corretagem segundo a tabela BOVESPA**

<b>Valor Negociado</b>	<b>Percentual do valor negociado</b>	<b>Valor Fixo</b>
<b>Até R\$ 135,07</b>	<b>-</b>	<b>R\$ 2,70</b>
<b>De R\$ 135,08 a R\$ 498,62</b>	<b>2,0%</b>	<b>-</b>
<b>De R\$ 498,63 a R\$ 1514,69</b>	<b>1,5%</b>	<b>R\$ 10,06</b>
<b>Acima de R\$ 3029,39</b>	<b>0,5%</b>	<b>R\$ 25,21</b>

Fonte: <http://veja.abril.com.br/especiais/digital4/tabelas.html>

**Tabela 2 - Emolumento cobrado de operações no mercado acionário**

<i>Tipo de operação</i>	<i>Taxa de negociação</i>	<i>Taxa de liquidação</i>	<i>Total</i>
Operação normal	0,027%	0,008%	0,035%
Operação <i>day trade</i>	0,019%	0,006%	0,025%

Fonte dos dados: <http://www.bmfbovespa.com.br/pt-br/regulacao/acoes/custos-operacionais/custos-operacionais.aspx?idioma=pt-br>

O investidor paga um valor mensal relativo à custódia de ações pela CBLC. Se o investidor possui ações em carteira, ele paga um valor mensal pela custódia destes títulos feita pela CBLC (veja Tabela 3). Caso não seja proprietário de nenhuma ação durante o mês o valor de custódia é reduzido. Segundo a BMF&BOVESPA este valor é cobrado da corretora e pode ser repassado ao investidor. Por este motivo, este custo pode variar de corretora para corretora.

O investidor deve pagar também ao final do mês o imposto de renda sobre o lucro líquido de cada operação de venda realizada neste período. Dependendo do tipo de negociação que o investidor realizou a porcentagem do imposto pode aumentar. Caso tenha

Tabela 3 - Custos de Manutenção de Contas de Custódia.

<i>Descrição</i>	<i>Valor mensal</i>
Conta sem movimentação ou posição	R\$3,00
Conta com movimentação ou posição de ativos de renda variável	R\$6,90

Fonte: <http://www.bmfbovespa.com.br/pt-br/regulacao/custos-e-tributos/custos-operacionais/acoes.aspx?Idioma=pt-br>

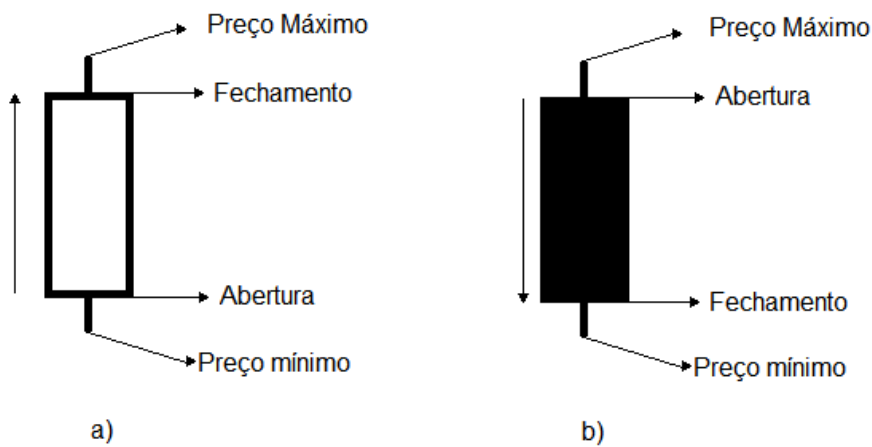
feito operações *day trade*, então o investidor precisa pagar 20% do lucro líquido obtido em cada operação *day trade*. Por outro lado, se o investidor realizou apenas operações normais, ou seja, se realizou a compra de um ativo em um dia e a venda em qualquer outro dia subsequente a este, então ele deverá pagar 15% do lucro líquido em todas as operação de venda normal. Contudo, se a soma do valor de cada negociação feita no mês for menor que 20 mil reais apenas as operação *day trade* serão cobradas do investidor.

## 2.4 Outros conceitos de mercado

Para o resto deste trabalho é necessário introduzir a noção de *candles* e *drawdown* frequentemente utilizados na análise técnica de mercado.

### 2.4.1 Candle

Entenda-se como um *candle* uma unidade que agrupa dados de um intervalo de tempo, apresentando os preços de abertura, fechamento, máximo e mínimo deste intervalo. Um *candle* pode ser formado por qualquer intervalo de tempo, seja de dez em dez minutos, de hora em hora ou até mesmo de dia em dia. O *candle* preenchido de branco indica que o período fechou em alta, enquanto o *candle* preenchido em preto indica que o período fechou em baixa (veja **Erro! Fonte de referência não encontrada.**). Estas cores podem variar de sistema para sistema.



**Figura 1 - Representação de um candle.**

O *candle* (a), sem preenchimento, indica um movimento ascendente do preço durante o período. O *candle* (b), preenchido, indica um movimento descendente no preço.

A Figura 2 mostra uma variação do gráfico de *candles*, chamado de gráfico *candlestick*. Neste caso, o vermelho indica período de queda no preço e o verde indica período de alta.

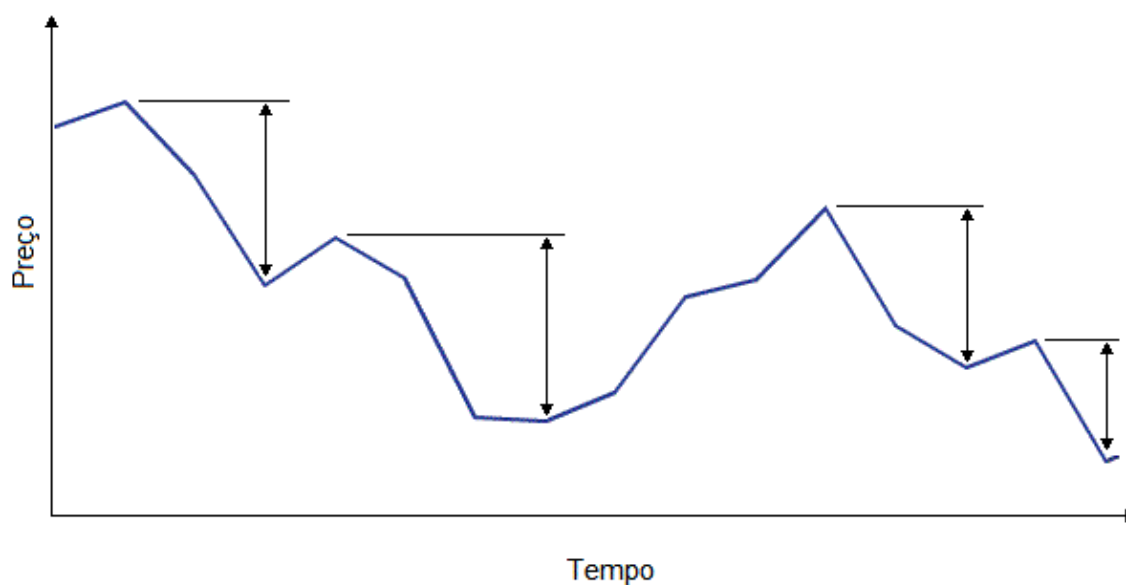


**Figura 2 - Gráfico *candlestick* do preço da ação PETR4.SA. Cada *candle* tem período de 1 dia.**

Fonte: <http://finance.yahoo.com/echarts?s=PETR4.SA+Interactive#symbol=PETR4.SA>

## 2.4.2 Drawdown

Segundo Sornette (2004), um *drawdown* é um decréscimo persistente no preço de um ativo durante dias consecutivos. Portanto, trata-se da perda acumulada do último máximo até o próximo mínimo do preço (veja Figura 3).



**Figura 3 - A figura mostra quatro *drawdowns* baseado em Sornette(2004), pág 52 .**

*Drawdowns* são indicadores importantes: eles medem diretamente a perda acumulada que um investimento pode sofrer. Também, quantificam o cenário de pior caso de um investidor comprando em um máximo local e vendendo no próximo mínimo.

Neste trabalho, o conceito de *drawdown* máximo é aplicado à maior desvalorização da carteira do agente financeiro ao se utilizar uma estratégia de investimento.

## 2.5 Sumário do Capítulo

Neste capítulo, apresentou-se uma visão geral sobre o domínio do problema trabalhado no projeto. Foram descritos os modelos de análise de mercado, os custos envolvidos para operar no mercado acionário e conceitos de mercado que serão utilizados no restante deste texto.

No próximo capítulo serão descritas as técnicas de otimização utilizadas no desenvolvimento deste projeto.

# 3

## Algoritmos de otimização

*Após uma introdução sobre algoritmos de otimização (Seção 3.1), este capítulo mostra o referencial teórico relativo à estrutura de dados de Árvore de decisão (Seção 3.2) e aos métodos de busca de soluções testados neste problema: Programação Genética (Seção 3.3) e Hill-Climbing (Seção 3.4).*

### 3.1 Introdução

Segundo Antoniou e Lu (2007), o processo de otimização é o processo de se obter o 'melhor', se for possível medir e mudar o que é 'bom' ou 'ruim'. Na prática, deseja-se maximizar ou minimizar uma função. Portanto a palavra 'ótimo' é usada para significar 'máximo' ou 'mínimo' dependendo das circunstâncias; 'ótimo' é um termo técnico que implica medida quantitativa e é uma palavra mais forte que 'melhor'.

Ainda segundo Antoniou e Lu (2007), a teoria da otimização é um ramo da matemática, que engloba o estudo quantitativos dos ótimos e métodos para encontrá-los. A prática de otimização, por outro lado, é a coleção de técnicas, métodos e procedimentos e algoritmos que podem ser usados para encontrar os ótimos.

Este capítulo descreve a estrutura utilizada na representação de soluções do problema e os algoritmos usados no processo de otimização.

### 3.2 Árvore de decisão

Inferência indutiva é o processo de transformar exemplos concretos para modelos genéricos. Seu objetivo é aprender como classificar objetos ou situações analisando um

conjunto de instâncias cujas classes são conhecidas. Classes são rótulos mutuamente exclusivos como, por exemplo, diagnósticos médicos, categoria de imagens ou modos de falha. Instâncias são tipicamente representadas como vetores de atributo-valor. A entrada para o aprendizado consiste em um conjunto destes vetores, cada um pertencente a uma classe conhecida, e o resultado consiste de um mapeamento de valores de atributos para classes. Este mapeamento deve classificar com precisão tanto instâncias conhecidas como outras instâncias desconhecidas (QUINLAN,1996).

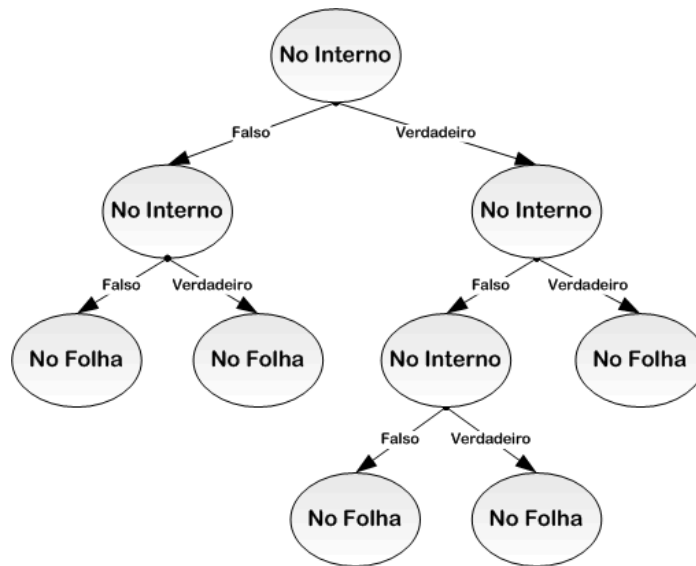


Figura 4 - Árvore de decisão binária

Uma árvore de decisão é um formalismo para expressar esse mapeamento. Uma árvore é um nó folha rotulado com uma classe ou uma estrutura consistindo de um nó de teste ligado a duas ou mais subárvores. Um nó de teste calcula um resultado baseado nos atributos de uma instância, no qual cada possível resultado é associado com uma das subárvores. Uma instância é classificada começando pelo nó raiz da árvore. Se esse nó é um teste, o resultado para a instância é determinado e o processo continua usando a subárvore apropriada. Quando uma folha é eventualmente encontrada, seu rótulo fornece a classe prevista para a instância. (QUINLAN, 1996).



### 3.3 Programação genética

Os algoritmos evolucionários simulam o processo de evolução Darwiniano e operações genéticas que ocorrem naturalmente nos cromossomos (KOZA, 1992). São algoritmos probabilísticos que fornecem um mecanismo de busca paralela e adaptativa baseado no princípio de sobrevivência dos mais aptos e ocorrência de mutações. Atuam como o proposto na teoria de evolução de Darwin, em que o processo de seleção privilegia os indivíduos que melhor se adaptam a um determinado ambiente. Estes indivíduos possuem maiores chances de se reproduzirem e assim passar seu código genético para seus descendentes. Durante essa reprodução pode também ocorrer outras formas de mudanças no código genético de um indivíduo chamadas mutação. O esperado após vários ciclos de evolução é a presença de indivíduos mais adaptados a um determinado ambiente.

Um ciclo de evolução pode ser descrito como a seleção de pais (indivíduos pertencentes à população) para o "cruzamento", seguida da recombinação de seus genes e mutações, dando origem assim a novos indivíduos, a prole. Estes por sua vez formarão a nova população.

Algoritmos evolucionários podem ser usados para otimizar a solução de um problema (por exemplo, encontrar um indivíduo que se adapta melhor as condições dadas) tendo em mãos as condições as quais uma população de indivíduos é exposta.

Uma das técnicas de computação evolucionária é a programação genética originada na década de 1980. Esta técnica procura fornecer um método para a criação automática de indivíduos a partir de uma descrição em alto nível do problema a ser atacado (REZENDE, 2003). Em programação genética usam-se árvores sintáticas para representar indivíduos, sendo árvore de decisão um tipo específico a ser utilizado.

Na programação genética, populações de centenas de milhares de indivíduos são geneticamente gerados. Essa criação é feita usando o princípio de Darwin da sobrevivência e reprodução do mais adaptado juntamente com operações de recombinação genética apropriadas para o "acasalamento" de indivíduos. Um indivíduo que resolve (ou resolve

aproximadamente) um dado problema pode emergir desta combinação de seleção natural Darwiniana e operações genéticas (KOZA, 1992).

Cada indivíduo na população é avaliado em termos de quão bem ele reage a um problema particular do ambiente. Essa avaliação é chamada de função objetivo. A natureza da função objetivo varia com o problema (KOZA, 1992).

Segundo Eiben (2003), os componentes mais importantes de um algoritmo evolucionário são:

- Representação (definição dos indivíduos);
- Função objetivo (ou função *fitness*) – que avalia a qualidade de um indivíduo;
- População;
- Mecanismo de seleção de pais;
- Operadores de variação, recombinação e mutação;
- Mecanismos de seleção do indivíduo sobrevivente.

### **3.3.1 Representação do indivíduo**

A idéia geral em programação genética é usar árvores sintáticas para descrever indivíduos. Estas árvores capturam expressões em uma sintaxe formal que podem ser expressões aritméticas, fórmulas lógicas ou códigos escritos em linguagem de programação (EIBEN, 2003).

### **3.3.2 Função objetivo**

A função objetivo, ou função *fitness*, é a força impulsionadora na seleção natural Darwiniana e da mesma forma na programação genética. Na natureza, o *fitness* de um indivíduo é a probabilidade de que ele sobreviva à idade de reprodução e reproduza. Essa medida pode ser ponderada para considerar a quantidade de sua prole. No mundo artificial dos algoritmos matemáticos, mede-se o *fitness* de uma forma e então usamos essa medida

para controlar a aplicação de operações que modificam as estruturas na população artificial.

A abordagem mais comum para medir o *fitness* é criar uma fórmula explícita para cada indivíduo da população (KOZA, 1992). Esta medida é conhecida como função objetivo.

### **3.3.3 População**

A população é o conjunto dos indivíduos e seu é manter possíveis soluções para o problema. A população forma a unidade da evolução. Indivíduos são objetos estáticos que não mudam ou se adaptam. Estas mudanças só ocorrem na população. Dado a representação de um indivíduo, definir uma população pode ser simples como especificar quantos indivíduos ela pode possuir (EIBEN, 2003).

### **3.3.4 Mecanismos de seleção de pais e sobreviventes**

Segundo Eiben (2003), o papel da seleção de pais é para distinguir entre os indivíduos baseado em sua qualidade, em particular, para permitir que os melhores indivíduos se tornem pais na próxima geração. Neste contexto, indivíduos de alta qualidade têm uma chance maior de se tornarem pais, enquanto que os de baixa qualidade têm chance menor.

O mecanismo de seleção de sobreviventes é chamado após os filhos dos pais selecionados terem sido criados. Alguns filhos gerados são escolhidos para a próxima geração. A decisão é usualmente baseada em seus valores de *fitness*. O objetivo deste processo, juntamente com o mecanismo de seleção de pais, é aplicar melhorias de qualidade na população.

### **3.3.5 Recombinação genética**

O papel dos operadores de recombinação é criar novos indivíduos a partir daqueles já presentes na população. Este processo gera novas soluções candidatas para o problema proposto (EIBEN, 2003).

Segundo Koza (1992) os dois principais operadores pra programação genética são a reprodução e o *crossover*. Existe também a possibilidade fazer uso de operações secundárias como a mutação.

#### **3.3.5.1 Crossover**

A operação de *crossover* (cruzamento, recombinação sexual) para programação genética cria uma variação na população produzindo novos indivíduos que possuem características de cada indivíduo pai. A operação começa com dois indivíduos pais e gera dois novos indivíduos (KOZA, 1992). A operação começa selecionando um ponto aleatório em cada pai para ser o ponto de *crossover* para aquele indivíduo. Os fragmentos resultantes desta etapa são então trocados gerando os indivíduos filhos.

#### **3.3.5.2 Reprodução**

A operação de reprodução para a programação genética é o mecanismo básico da seleção natural Darwiniano e a sobrevivência do mais apto. A operação de reprodução é assexuada e trabalha em somente um indivíduo pai e produz somente um indivíduo filho em cada vez que é realizada.

A operação de reprodução consiste em duas etapas. Primeiro um indivíduo é selecionado de acordo com algum método de seleção baseado na função *fitness*. Segundo, o indivíduo é copiado, sem alteração, da população atual para a nova população.

### 3.3.6 Mutação

A operação de mutação introduz mudanças aleatórias em indivíduos. Em algoritmos genéticos convencionais a operação de mutação pode ser benéfica ao reintroduzir diversidade na população que pode tender a convergir prematuramente. Mutação é assexuada e opera apenas em um indivíduo. A mutação inicia selecionando um nó aleatório no indivíduo. A operação então remove este nó e todos os nós descendentes do mesmo e insere uma nova sub-árvore gerada aleatoriamente.

### 3.3.7 Algoritmo de otimização evolucionário

O esquema geral de um algoritmo evolucionário descrito por (EIBEN, 2003) é apresentado em Algoritmo 1.

---

INÍCIO

    INICIALE população com soluções candidatas  
    AVALIE cada candidato  
    REPITA ATE (CONDIÇÃO DE TERMINO for satisfeita )  
        SELECIONE pais  
        RECOMBINE pares de pais  
        APLIQUE MUTAÇÃO na prole resultante  
        AVALIE novos candidatos  
        SELECIONE indivíduos para a próxima geração  
    FIM REPITA

FIM

---

**Algoritmo 1 - Esquema geral de um algoritmo evolucionário (EIBEN, 2003) .**

## 3.4 Hill-Climbing

Segundo Voudouris e Tsang (2001), existem problemas que devido à grande ou ilimitada quantidade de soluções são difíceis de resolver usando métodos completos. Uma das maneiras de achar soluções para tais problemas é sacrificar a checagem de todo o espaço de busca e usar heurísticas para definir um subconjunto de soluções que serão testadas. Alguns dos melhores métodos conhecidos que utilizam esta estratégia são os métodos de busca local, a forma básica do que é referido como *hill-climbing* (veja Algoritmo 2).

A execução do *hill-climbing* inicia sua busca com uma solução candidata, que pode ser aleatória ou heurísticamente gerada. A busca se move para um vizinho que é melhor de acordo com a função objetivo e termina se nenhum vizinho melhor puder ser encontrado, ou se um critério de parada for atingido. Todo o processo pode ser repetido de diferentes pontos de início.

Ainda segundo Voudouris e Tsang (2001), para realizar o *hill-climbing*, é preciso definir o seguinte:

- a) Representação para a solução candidata.
- b) Uma função objetivo: dado qualquer solução candidata, esta função retorna um valor numérico. O problema é visto como um problema de otimização de acordo com a função objetivo (a qual deve ser minimizada ou maximizada).
- c) Uma função de vizinhança que mapeia cada solução candidata  $x$  (frequentemente chamada de *estado*) para um conjunto de outras soluções candidatas (que são chamadas de soluções vizinhas de  $x$ ). Em outras palavras, o domínio de uma função de vizinhança é um conjunto de todos os candidatos a solução,  $S$ . Seu alcance é o conjunto potência de  $S$ .

---

*Função SUBIDA\_DE\_ENCOSTA(problema) retorna um estado que é um máximo local*

---

**Entradas:** problema, um problema

**Variáveis locais:** corrente, um nó

Vizinho, um nó

Corrente  $\leftarrow$  CRIAR\_NO (ESTADO\_INICIAL [problema])

**Repita**

Vizinho  $\leftarrow$  um sucessor de corrente com valor mais alto

Se VALOR[vizinho]  $\leq$  VALOR[corrente] então retornar ESTADO[corrente]

Corrente  $\leftarrow$  vizinho

---

**Algoritmo 2 - Algoritmo de *hill-climbing* extraído de Russel e Norvig (2004).**

Neste trabalho, a representação da solução candidata é cada agente financeiro encontrado na busca (definido na Seção 4.5.1). A função objetivo, ou função de avaliação, é a aquela definida na Seção 4.5.2 e as funções de vizinhança são as funções de variação definidas na Seção 4.5.3.

## 3.5 Sumário do capítulo

Este capítulo apresentou o referencial teórico sobre as técnicas de otimização levantadas para o desenvolvimento do trabalho. Também foi apresentado o referencial teórico sobre a estrutura de árvore binária escolhida para a representação das regras técnicas no *software*. A escolha do algoritmo apropriado para o trabalho é apresentada e justificada no próximo capítulo, bem como as definições relativas ao projeto.





# 4

## Projeto

*Este capítulo apresenta as fases envolvidas na realização do projeto: Metodologia (Seção 4.1), Definição da hipótese (Seção 4.2), Definição e obtenção das bases de dados (Seção 4.3), Processamento das bases de dados (Seção 4.4), Definição das estruturas do programa (Seção 4.5) e Definição do algoritmo (Seção 4.6).*

### 4.1 Metodologia

A pesquisa em questão é empírica e experimental. Os tópicos de importância no desenvolvimento da pesquisa foram:

1. Definição da hipótese;
2. Definição e obtenção das bases de dados;
3. Processamento das bases de dados;
4. Definição das estruturas do programa;
5. Definição do algoritmo.

### 4.2 Definição da hipótese

A hipótese  $H_0$  que se pretende corroborar é definida como: “É possível encontrar agentes financeiros artificiais que descubrem regras técnicas lucrativas mensalmente no mercado acionário da BOVESPA utilizando busca local com o método *hill-climbing* e com estruturas de árvores de decisão”.

### 4.3 Definição e obtenção das bases de dados

Estudos relacionados à geração de regras técnicas em mercado acionário estrangeiro, como Lee (2009), Potvin (2004) e Schulenburg (1999, 2000, 2001), freqüentemente utilizam dados diários das ações. Neste trabalho, optou-se por usar dados *intraday* das ações para aumentar a precisão das regras encontradas. Outro fator motivador para a escolha de dados *intraday*, é a existência de leilões realizados no início de cada dia para se definir o valor das ações.

As bases de dados utilizadas pelo programa foram cedidas pela empresa SI2 (<http://si2.inf.br/>). Cada arquivo contém os dados históricos de uma única ação (veja Tabela 4). Cada linha em um arquivo contém dados, separados por ponto-e-vírgula, relativos a um único minuto.

**Tabela 4 - Estrutura do arquivo com dados da PETR4.**

<i>Ativo</i>	<i>Data</i>	<i>Horário</i>	<i>Abert.</i>	<i>Máx.</i>	<i>Mín.</i>	<i>Fech.</i>	<i>Vol.</i>	<i>Qtd.</i>
PETR4	24/06/2008	10:02:00	40,46	40,54	40,36	40,54	5144486,00	117900
PETR4	24/06/2008	10:03:00	40,40	40,50	40,40	40,50	209660,00	4800
PETR4	24/06/2008	10:04:00	40,50	40,53	40,46	40,50	1883550,00	43100
PETR4	24/06/2008	10:05:00	40,51	40,51	40,46	40,46	388754,00	8900
...	...	...	...	...	...	...	...	...
PETR4	17/05/2010	15:24:00	29,66	29,66	29,66	29,66	121606,00	4100

Os arquivos contendo dados históricos das empresas Petrobrás (PETR4), Vale do Rio Doce (VALE5) e Banco do Bradesco (BBDC4) foram selecionados arbitrariamente como fonte de dados para o algoritmo (veja Tabela 5).

**Tabela 5 – Período total de cada base de dados**

<i>Empresa</i>	<i>Papel</i>	<i>Início</i>	<i>Fim</i>
Banco do Bradesco	BBDC4	24/06/2008 10:02:00	17/05/2010 15:23:00
Petrobrás	PETR4	24/06/2008 10:02:00	17/05/2010 15:24:00
Vale do Rio Doce	VALE5	24/06/2008 10:03:00	17/05/2010 15:24:00

## 4.4 Processamento das bases de dados

Devido à grande quantidade de dados contida em cada arquivo, foi necessário um pré-processamento das bases para agrupar os dados de 10 em 10 minutos. Após o agrupamento, cada linha dos arquivos representou dados de um período de 10 minutos. O agrupamento seguiu o processo descrito abaixo (veja Tabela 6):

- Ativo: recebe o nome do ativo, que se repete em todas as linhas do arquivo de dados por minutos.
- Data: recebe a data do último *candle* de minuto dos 10 *candles* de 1 minuto sendo processados.
- Horário: recebe o horário do último *candle* de minuto também.
- Abertura: recebe o valor de abertura do primeiro *candle* de minuto.
- Máximo: recebe o maior valor do campo ‘máximo’ das 10 linhas da base de minutos.
- Mínimo: recebe o menor valor do campo ‘mínimo’ das 10 linhas da base de minutos.
- Fechamento: recebe o valor de fechamento do último *candle* do conjunto de *candles* de 1 minuto.
- Volume negociado: recebe a soma dos volumes de todos os *candles* de 1 minuto.

- Quantidade: recebe a soma da quantidade negociada de todos os *candles* de 1 minuto.

**Tabela 6 – Processo de redução de candles. Reduz-se 10 *candles* de 1 minuto para 1 *candle* de 10 minutos, no arquivo da PETR4. A última linha da tabela é o resultado de um processamento parcial.**

<i>Ativo</i>	<i>Data</i>	<i>Horário</i>	<i>Abert.</i>	<i>Máx.</i>	<i>Mín.</i>	<i>Fech.</i>	<i>Vol.</i>	<i>Qtd.</i>
PETR4	24/06/2008	10:02:00	40,46	40,54	40,36	40,54	5144486	117900
PETR4	24/06/2008	10:03:00	40,4	40,5	40,4	40,5	209660	4800
PETR4	24/06/2008	10:04:00	40,5	40,53	40,46	40,5	1883550	43100
PETR4	24/06/2008	10:05:00	40,51	40,51	40,46	40,46	388754	8900
PETR4	24/06/2008	10:06:00	40,46	40,47	40,42	40,45	877361	20100
PETR4	24/06/2008	10:07:00	40,43	40,5	40,43	40,5	515216	11800
PETR4	24/06/2008	10:08:00	40,47	40,53	40,47	40,53	572546	13100
PETR4	24/06/2008	10:09:00	40,53	40,5	40,5	40,5	7319820	167500
PETR4	24/06/2008	10:10:00	40,51	40,59	40,51	40,59	1767510	40400
PETR4	24/06/2008	10:11:00	40,59	40,66	40,59	40,66	815818	18600
PETR4	24/06/2008	10:11:00	40,46	40,66	40,36	40,66	19494721	446200

## 4.5 Definição das estruturas do programa

### 4.5.1 Indivíduo

Um indivíduo da população será nomeado aqui como um agente financeiro artificial ou, por simplicidade, um agente. Todo agente neste trabalho é composto de duas árvores de decisões: uma para decisões de compra de ações e outra para decisões de venda. Ambas as árvores de decisões podem também retornar uma decisão de espera, se em sua execução atingir um nó folha com valor **falso**.

O modelo das árvores de decisões segue uma abordagem semelhante à estrutura usada por Kampouridis (2011), chamada de Árvores Genéticas de Decisão (*Genetic Decision Trees*). No modelo proposto aqui, todo nó da árvore ao ser processado, retorna um valor lógico. Cada nó interno da árvore é composto de uma expressão relacional e cada nó folha é uma constante lógica, **verdadeiro** ou **falso** (veja Figura 5).

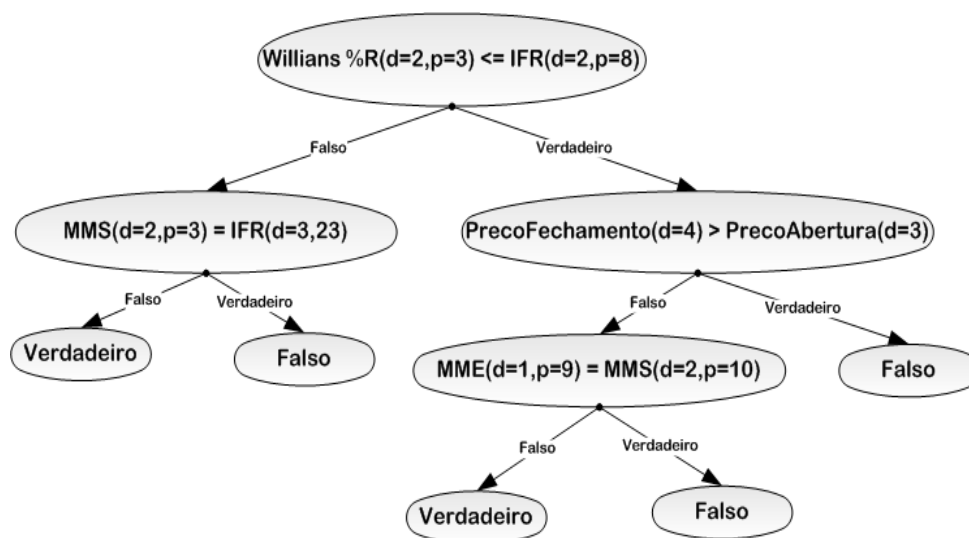


Figura 5 - Representação de uma possível árvore de compra ou de venda.

As árvores de decisão dos agentes financeiros são geradas da gramática livre de contexto  $G = \{N, T, R, P\}$  em que  $N$  é o conjunto de símbolos não terminais,  $T$  o conjunto de símbolos terminais,  $R$  o conjunto de regras e  $P$  o símbolo de partida:

$$N = \{ \langle RAIZ \rangle, \langle INTERNO \rangle, \langle EXP\_REL \rangle, \langle OP\_REL \rangle, \langle NO \rangle, \langle FOLHA \rangle, \langle EXP\_A \rangle \}$$

$$T = \{ CONST(), PABERT(d), PMAX(d, p), PMIN(d, p), PFECH(d), VABERT(d), VMAX(d, p), VMIN(d, p), VFECH(d), MMS(d, p), MME(d, p), IFR(d, p), UBB(d, p), LBB(d, p), MACD(d, p1, p2), PROC(d, p), WILR(d, p), OSCE(d, p), <, >, =, \neq, \leq, \geq, verdadeiro, falso \}$$

$$P = \{ \langle RAIZ \rangle \}$$

O conjunto de regras  $R$  foi definido na Tabela 7.

**Tabela 7 - Conjunto de regras da gramática G**

---

$\langle RAIZ \rangle \rightarrow \langle INTERNO \rangle$
$\langle INTERNO \rangle \rightarrow \text{IF } \langle EXP\_REL \rangle \text{ THEN } \langle NO \rangle \text{ ELSE } \langle NO \rangle$
$\langle EXP\_REL \rangle \rightarrow \langle EXP\_A \rangle \langle OP\_REL \rangle \langle EXP\_A \rangle$
$\langle OP\_REL \rangle \rightarrow \langle   \rangle   =   \neq   \leq   \geq$
$\langle NO \rangle \rightarrow \langle INTERNO \rangle   \langle FOLHA \rangle$
$\langle FOLHA \rangle \rightarrow verdadeiro   falso$
$\langle EXP\_A \rangle \rightarrow CONST(), PABERT(d), PMAX(d, p), PMIN(d, p), PFECH(d),$
$VABERT(d), VMAX(d, p), VMIN(d, p), VFECH(d), MMS(d, p), MME(d, p), IFR(d, p),$
$UBB(d, p), LBB(d, p), MACD(d, p1, p2), PROC(d, p), WILR(d, p), OSCE(d, p)$

---

É importante notar que o não terminal  $\langle EXP\_A \rangle$  leva a um conjunto de funções do domínio (veja Tabela 8) que são parametrizadas com um deslocamento  $d$  e/ou um período  $p$  ( $p1$  e  $p2$  também são períodos). O parâmetro  $d$  desloca  $d$  *candles* atrás, a partir do *candle* atual na execução, para calcular uma função. O período  $p$  indica que a partir do instante após o deslocamento, precisam-se reunir os  $p$  últimos *candles* para calcular uma dada função.

Definiu-se arbitrariamente que os valores de  $d$  e  $p$ , parâmetros das funções, devem estar contidos, respectivamente, nos intervalos  $[0, 6]$  e  $[0,30]$ . Os cálculos das funções de 10 a 18 apresentadas na Tabela 8 foram extraídos de Achelis (2010).

Ao criar um agente financeiro artificial na população, a primeira negociação que o mesmo realizará no mercado é de compra. Neste caso, o agente utilizará a árvore de compra para tomar suas decisões. Em um dado instante no mercado, ao percorrer a árvore de compra, se o caminho da árvore resultar em uma folha contendo o valor **verdadeiro**, a decisão da árvore será de utilizar todo o dinheiro disponível em carteira para comprar as ações de uma empresa pré-definida. Por outro lado, se o percurso da árvore resultar em um nó folha contendo o valor **falso**, então a árvore de compra retorna uma decisão de espera, significando que o agente financeiro artificial não realizará nenhuma operação.

O processo decisório de venda é análogo ao de compra, ao se percorrer a árvore de venda e encontrar um nó com valor **verdadeiro**, então a decisão é de venda, caso encontre um nó com valor **falso**, então o agente decide esperar.

Os agentes financeiros sempre realizam operações absolutas, ou seja, em uma operação de compra o agente utilizará todo o seu dinheiro em carteira para comprar ações da empresa em questão. Analogamente, em uma única operação de venda, o agente venderá todas as ações que possui em carteira. Neste trabalho, o agente financeiro artificial negocia apenas ações de uma única empresa pré-definida no início do algoritmo.

**Tabela 8 - Conjunto de funções usadas em <EXP\_A>.**

<i>Num.</i>	<i>Função</i>	<i>Descrição</i>
1	double CONST()	Função constante. Pode retornar valores no intervalo inteiro [0,100] ou no intervalo real [0,1]
2	double PABERT(int d)	Preço de abertura
3	double PMAX(int d, int p)	Preço máximo
4	double PMIN(int d, int p)	Preço mínimo
5	double PFECH(int d)	Preço de fechamento
6	double VABERT(int d)	Volume de negociação na abertura
7	double VMAX(int d, int p)	Volume de negociação máximo
8	double VMIN(int d, int p)	Volume de negociação mínimo
9	double VFECH(int d)	Volume de negociação no fechamento
10	double MMS (int d, int p)	Média móvel simples do preço
11	double MME(int d, int p)	Média móvel exponencial
12	double IFR(int d, int p)	Índice de Força Relativa
13	double UBB(int d, int p)	Linha superior do indicador <i>Bollinger Bands</i>
14	double LBB(int d, int p)	Linha inferior do indicador <i>Bollinger Bands</i>
15	double MACD(int d, int p1, int p2)	Indicador Média Móvel Convergência/Divergência
16	double PROC(int d, int p)	Indicador de taxa de variação do preço ( <i>Price Rate of Change</i> )
17	double WILR(int d, int p)	Indicador Williams %R.
18	double OSCE(int d, int p)	Oscilador Estocástico



## 4.5.2 Função objetivo

A qualidade de um agente financeiro artificial é definida pela função  $f$  e considera quatro pontos importantes:

1. Razão entre valor final de carteira após investimento por valor inicial.
2. *Drawdown* máximo.
3. Razão entre a quantidade de operações boas realizadas pela quantidade de operações totais.
4. Quantidade total  $n$  de operações realizadas.

As funções que representam cada um destes pontos são respectivamente:

$$(1) \quad L = \left( \frac{\text{valor}_{final}}{\text{valor}_{inicial}} \right)$$

$$(2) \quad D = (1 - \text{drawdown}_{maximo})$$

$$(3) \quad O_b = \frac{2 * \text{vendas\_boas}}{\text{operacoes}}$$

$$(4) \quad O_t(n) = \begin{cases} \frac{\text{operacoes}}{n}, & \text{operacoes} < n \\ 1, & n \geq \text{operacoes} \end{cases}$$

Finalmente, a função *fitness*  $f$  geral para avaliar a qualidade de um agente é dada por:

$$(5) \quad F(n) = \begin{cases} L + D + O_b + O_t(n), & \text{se } \text{operacoes} > 0 \\ 0, & \text{se } \text{operacoes} = 0 \end{cases}$$

Seguem os significados das variáveis e parâmetros:

- $\text{valor}_{final}$  é o capital do investidor no final do período de treinamento, considerando as ações, se houver.
- $\text{valor}_{inicial}$  é o capital inicial com o qual o agente inicia seu investimento. Neste trabalho, todo agente inicia com R\$20.000,00.
- $\text{drawdown}_{maximo}$  é a desvalorização máxima da carteira do agente em um dado momento no período de treinamento.

- *vendas\_boas* é a quantidade de vendas realizadas em que o preço da ação na venda foi maior que seu preço na respectiva operação de compra.
- *operacoes* é a quantidade total de operações realizadas pelo agente. Essencialmente, a soma da quantidade de operações de compra com a quantidade de operações de venda.
- *n* é o número mínimo de operações que o usuário quer que ocorra.

O papel da função *f* é avaliar a qualidade dos agentes financeiros no período de treinamento. No período de teste, esta não é utilizada tendo em vista que este período apenas executa o melhor agente encontrado no treinamento.

Esta função pode ser modificada, à medida que funções melhores forem descobertas para avaliação de soluções do domínio do problema.

### 4.5.3 Operações de variação

Por questões de otimização e para viabilizar os experimentos, optou-se por não usar população na Programação Genética. Devido à grande quantidade de dados *intraday* contidos em cada base (cerca de 19000), esta abordagem se tornou lenta e inviável para o projeto. Portanto não foram utilizadas as operações de *crossover* e reprodução descritas no Capítulo 3. Por não utilizar população, o método de busca de soluções se resume no método de otimização *hill-climbing*, evoluindo apenas um indivíduo por execução.

Note-se que as estruturas modificadas para otimizar a solução do problema são as árvores do agente financeiro e que a mutação é aplicada apenas nestas árvores caracteriza a o processo evolutivo deste indivíduo.

Os operadores de variação criados neste trabalho estendem a idéia de mutação do indivíduo descrita na Seção 3.3.6. Utilizamos mais dois tipos de variação no indivíduo além daquele já apresentado:

- **Variação 1:** escolhe-se um nó aleatoriamente na árvore. Se o nó for um nó interno, então se substitui a expressão relacional deste por uma outra

inteiramente nova gerada aleatoriamente. Se o nó for folha, então se substitui o valor dela pela negação deste valor.

- **Varição 2:** escolhe-se um nó aleatoriamente na árvore. Se o nó for folha, ocorre como na variação 1. Se o nó for interno, escolhe-se algum valor terminal da relação booleana e troca por outro terminal gerado aleatoriamente que possa ser substituído por este.

## 4.6 Definição do algoritmo

Neste trabalho, dois algoritmos foram selecionados para o processo de otimização: Programação Genética e *hill-climbing*. Os primeiros testes foram realizados utilizando o algoritmo de Programação Genética descrito na seção 3.3. Porém, este método apenas obteve bons resultados com grandes populações (veja detalhes na Tabela 9), o que resultou em um processamento demorado, inviabilizando a execução dos experimentos.

**Tabela 9 - Parâmetros para o algoritmo genético.**

<i>Representação</i>	<i>Indivíduo (descrito na seção 3.3.1)</i>
Recombinação	<i>Crossover</i>
Probabilidade de Recombinação	80%
Mutação	Varição 1
Probabilidade de Mutação	20%
Seleção de pais	Torneio com substituição ( $u = 50, k = 10$ )
Tamanho da População	300-500
Inicialização	Aleatória
Critério de termino	Número de gerações (50-100)

Portanto, fez-se necessário o uso de outro método de otimização: o *hill-climbing* (veja Seção 3.4). Este método conseguiu resultados semelhantes ao do método da Programação Genética, mas com a vantagem de realizar operações de variação em apenas um indivíduo desconsiderando a população do método anterior.

O aprendizado do agente é feito usando o conjunto de dados coletados. Estes dados referem-se a variações do preço de um ativo em um determinado período de tempo. Os dados pertencentes a este período são submetidos ao agente para que ele configure suas árvores de decisão. Este período é conhecido como conjunto de treinamento.

Após este treinamento, o agente é novamente submetido a outro conjunto de dados, ainda não vistos, conhecido como conjunto de teste. O desempenho do agente é então avaliado após este período.

Fazendo uso do *hill-climbing* com reinício aleatório (RUSSEL, NORVING 2004), conduziu-se uma série de buscas no espaço de soluções a partir de estados iniciais gerados ao acaso. Ou seja, para cada conjunto de treinamento e testes, conduziram-se várias instâncias do *hill-climbing*, adotando como agente final, o indivíduo que obteve melhor qualidade durante estas execuções, segundo a função objetivo  $f$  descrita neste trabalho.

Este agente final será usado para fins de comparação com outras estratégias comuns de mercado.

#### 4.6.1 Parâmetros do algoritmo

Diversos parâmetros influenciam as execuções do experimento. Estes podem ser dividido em 3 subconjuntos. Estes subconjuntos descrevem os parâmetros e os valores usados para em cada experimento (veja Tabela 10, Tabela 11, Tabela 12).

**Tabela 10 - Parâmetros para função objetivo**

<i>Parâmetro</i>	<i>Valor</i>
Quantidade mínima de operações	40

A quantidade mínima de operações é definida como 20 vezes a quantidade de meses do período de treinamento. Esta quantidade mínima foi definida, levando-se em conta que o mês tem 22 dias de negociações. Portanto, deseja-se em média uma operação por dia.

**Tabela 11 - Parâmetros para agente financeiro**

<i>Parâmetro</i>	<i>Valor</i>
Capital Inicial	R\$ 20.000,00
Período de treinamento	2 meses
Período de teste	1 mês
Probabilidade de gerar operação	60%
Período máximo	30 instantes
Deslocamento máximo	5 instantes
Taxas	As mesmas da seção 2.3 exceto CBLC

A probabilidade de gerar operações refere-se à probabilidade de cada folha da árvore de decisão assumir o valor verdadeiro, no processo de geração das árvores aleatórias do agente.

**Tabela 12 - Parâmetros para *hill-climbing***

<i>Parâmetro</i>	<i>Valor</i>
Função objetivo	A mesma de Função objetivo 4.5.2
Mutações	As mesmas de Operações de variação 4.5.3
Número de Iterações	10.000

## **4.7 Sumário do capítulo**

Neste capítulo foram apresentados o formato da base de dados utilizada, seu conteúdo e o processamento para sua redução. Também foram mostradas a modelagem dos agentes financeiros artificiais e a técnica *Hill-Climbing* utilizada para a solução do problema.

Os resultados obtidos das definições realizadas neste capítulo e as comparações entre as estratégias de investimento são apresentados no próximo capítulo.

# 5

## Resultados

### 5.1 Considerações iniciais

Com base na descrição do algoritmo da Seção 4.6, neste capítulo é realizada a comparação dos resultados obtidos através do experimento realizado com estratégias clássicas de mercado como *buy and hold* e investimento de renda fixa.

A estratégia de *buy and hold* pode ser representada por um agente financeiro que realiza uma compra no início de um período e uma venda no fim do período, não realizando qualquer outra operação durante o período.

O investimento de renda fixa consiste na aplicação do dinheiro do agente em banco rendendo juros fixos mensais. A cada mês, o rendimento é de um determinado valor que é determinado pela taxa de juros. Atualmente, a taxa de juros cobre o intervalo de 0.5% a 0.7% ao mês. Neste estudo assumiu-se a taxa de juros como sendo 0.5% ao mês.

Para fins de comparação, somente os melhores agentes do período de teste são utilizados.

### 5.2 Resultados

Devido à extensão dos resultados, somente algumas instâncias do experimento serão descritas. A escolha destas instâncias foi feita aleatoriamente.

## 5.2.1 Ações PETR4.SA

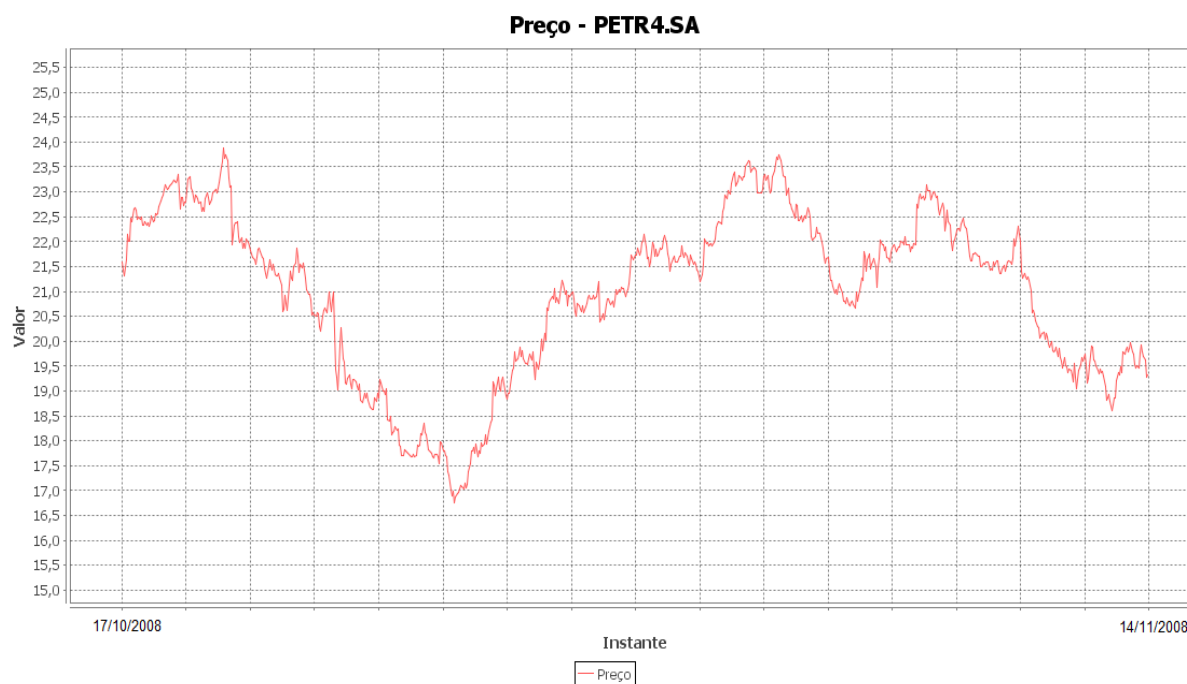
Esta seção mostra os resultados dos agentes obtidos para as ações da Petrobrás.

### 5.2.1.1 Período de teste número 1

As características do período de teste número 1 é descrito na Tabela 13. O gráfico para este período é representado na Figura 6.

**Tabela 13 - Descrição do período de teste número 1.**

<i>Ação</i>	<i>Início do período</i>	<i>Fim do período</i>	<i>Preço de abertura</i>	<i>Preço de fechamento</i>	<i>Varição</i>
PETR4.SA	17/10/2008	14/11/2008	21,59	19,32	-2,27



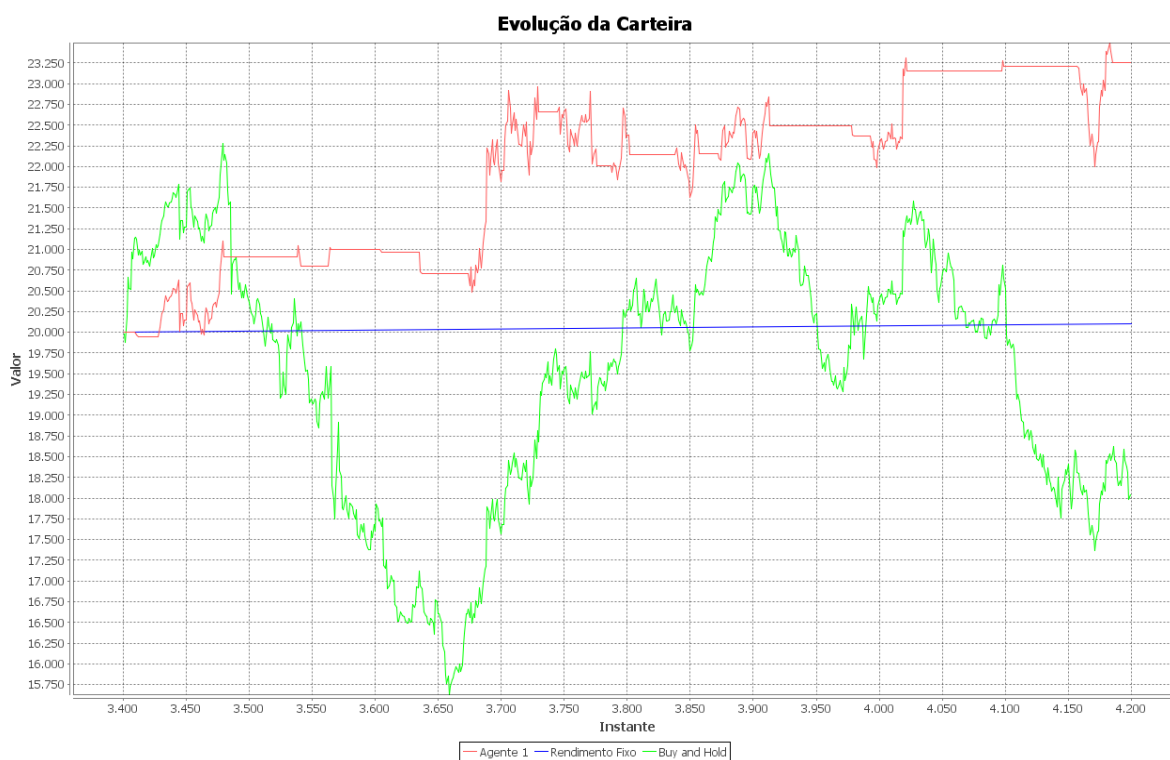
**Figura 6 - Gráfico do preço do período de teste número 1.**



Na Tabela 14 é mostrado o desempenho do melhor agente encontrado para o período número 1. Também está incluído o desempenho do agente de *buy and hold* e do investimento em banco. A Figura 7 representa graficamente este período, na qual cada instante representa o intervalo de 10 candles.

**Tabela 14 - Desempenho dos agentes do período 1.**

<i>Desempenho</i>	<i>Valor Final</i>	<i>Drawdown máximo</i>	<i>Quantidade de Operações</i>	<i>Lucro</i>
Agente Treinado	23256,00	5,8%	34	3.256,00
Investimento em banco	20100,00	0	0	100,00
<i>Buy and hold</i>	17867,00	12,23%	2	-2133,00



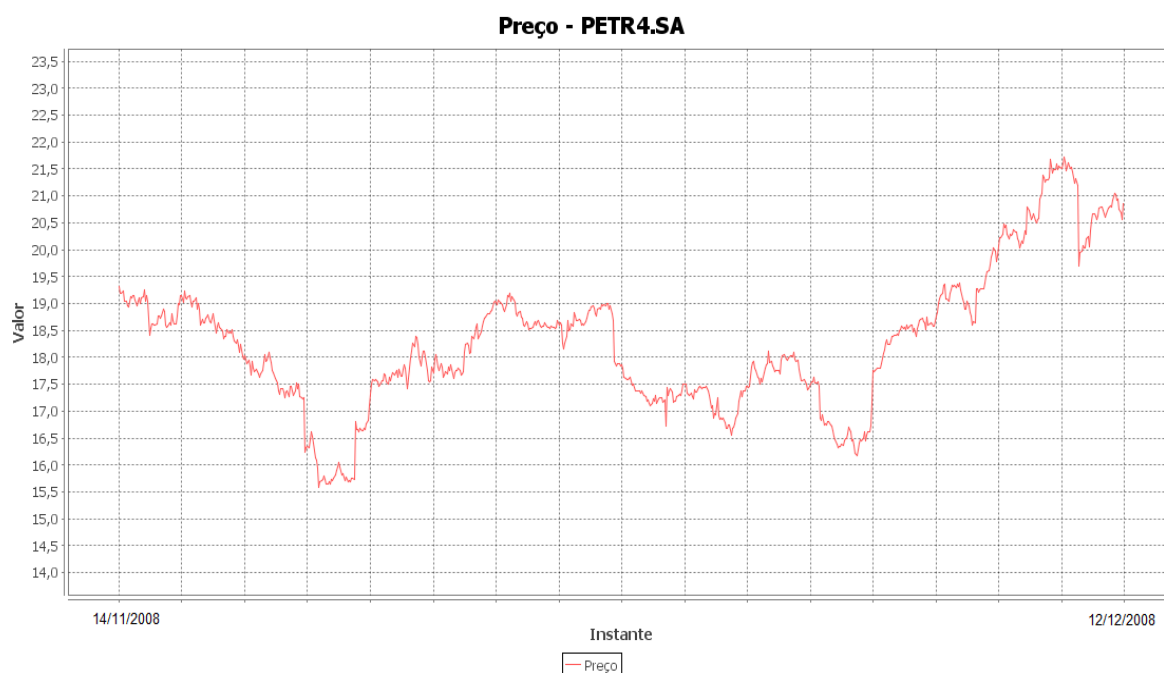
**Figura 7 - Gráfico de desempenho do período 1.**

### 5.2.1.2 Período de teste número 2

As características do período de teste número 2 é descrito na Tabela 15. O gráfico para este período é representado na Figura 8.

**Tabela 15 - Descrição do período de teste número 2.**

<i>Ação</i>	<i>Início do período</i>	<i>Fim do período</i>	<i>Preço de abertura</i>	<i>Preço de fechamento</i>	<i>Variação</i>
PETR4.SA	14/11/2008	12/12/2008	19,32	20,85	1,53

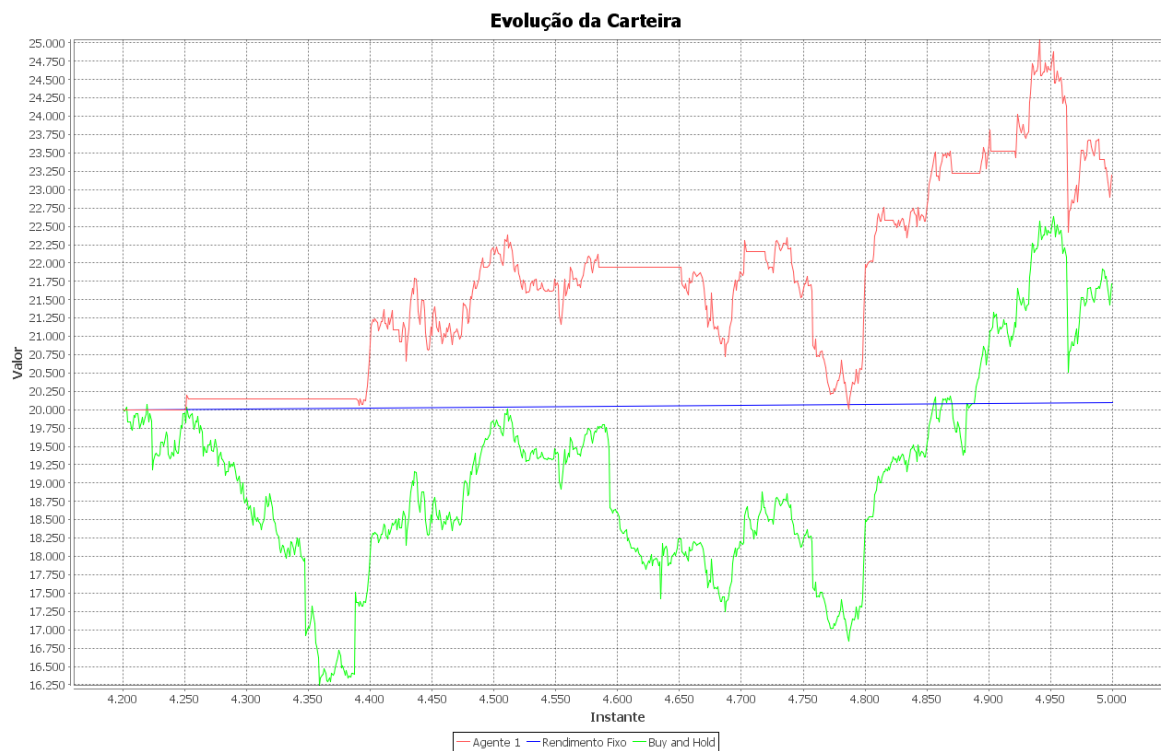


**Figura 8 - Gráfico do preço do período de teste número 2.**

Na Tabela 16 é mostrado o desempenho do melhor agente encontrado para o período número 2. Também está incluído o desempenho do agente de *buy and hold* e do investimento em banco. A Figura 9 representa graficamente este período.

**Tabela 16 - Desempenho dos agentes do período 2.**

<i>Desempenho</i>	<i>Valor Final</i>	<i>Drawdown</i>	<i>Quantidade de Operações</i>	<i>Lucro</i>
Agente Treinado	23.136,20	10,59%	31	3.136,20
Investimento em banco	20.100,00	0	0	100,00
Buy and hold	21.310	12,24%	2	1.310,00



**Figura 9 - Gráfico de desempenho do período 2.**

### 5.2.1.3 Período de teste número 3

As características do período de teste número 3 é descrito na Tabela 17 Tabela 16. O gráfico para este período é representado na Figura 10.

Tabela 17 - Descrição do período de teste número 3.

<i>Ação</i>	<i>Início do período</i>	<i>Fim do período</i>	<i>Preço de abertura</i>	<i>Preço de fechamento</i>	<i>Varição</i>
PETR4.SA	13/03/2009	09/04/2009	26,39	29,66	3,27

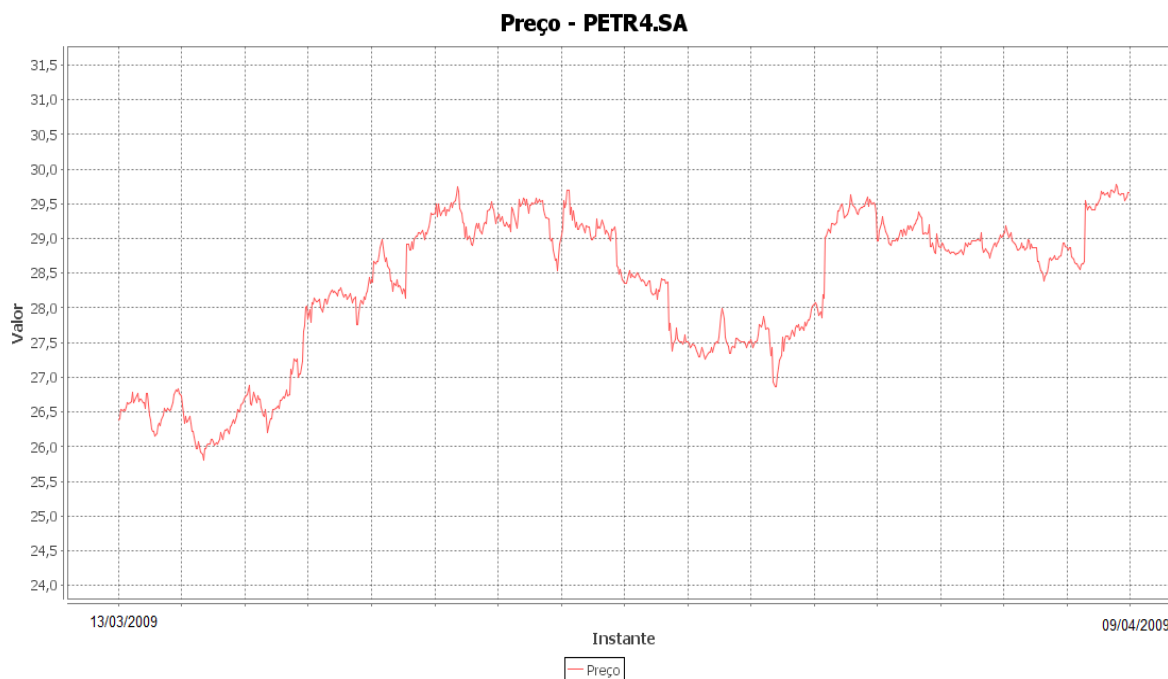
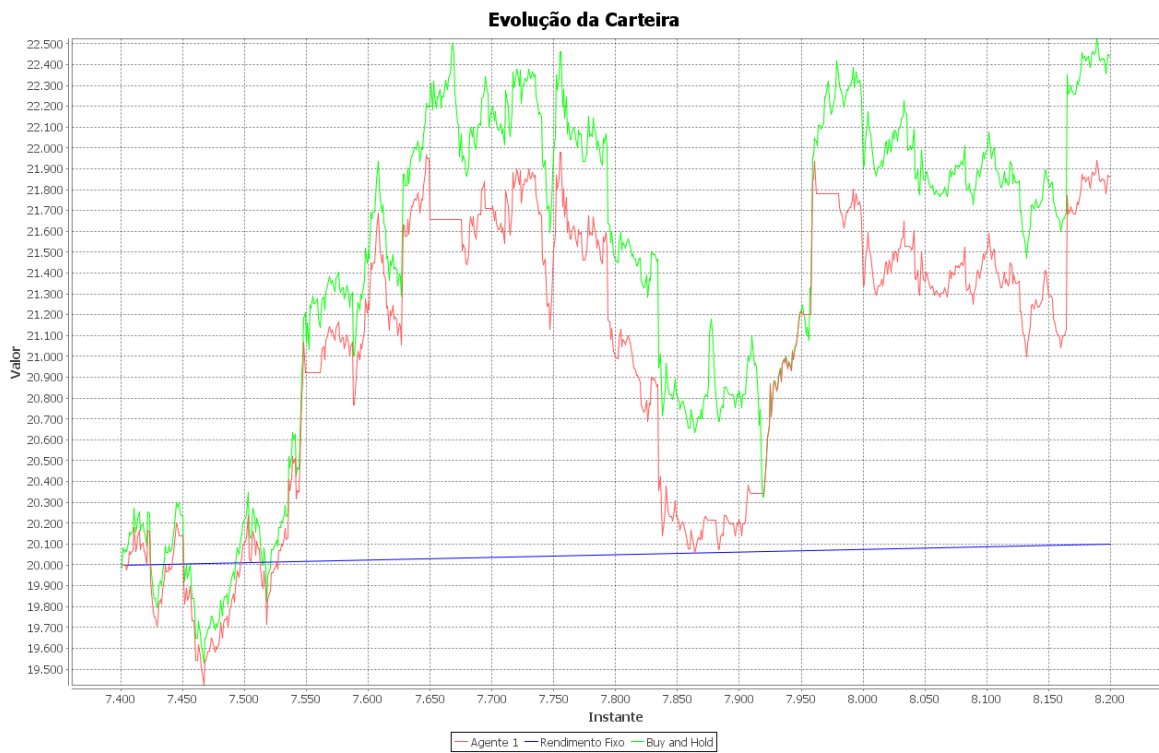


Figura 10 – Gráfico do preço do período de teste número 3.

Na Tabela 18 é mostrado o desempenho do melhor agente encontrado para o período número 3. Também está incluído o desempenho do agente de *buy and hold* e do investimento em banco. A Figura 11 representa graficamente este período.

**Tabela 18 - Desempenho dos agentes do período 3.**

<i>Desempenho</i>	<i>Valor Final</i>	<i>Drawdown</i>	<i>Quantidade de Operações</i>	<i>Lucro</i>
Agente Treinado	21.903,96	8,74%	25	1.903,96
Investimento em banco	20.100,00	0	0	100,00
Buy and hold	22.082,00	23,32%	2	2.082,00



**Figura 11 - Gráfico de desempenho do período 3.**

## 5.2.2 Ações VALE.SA

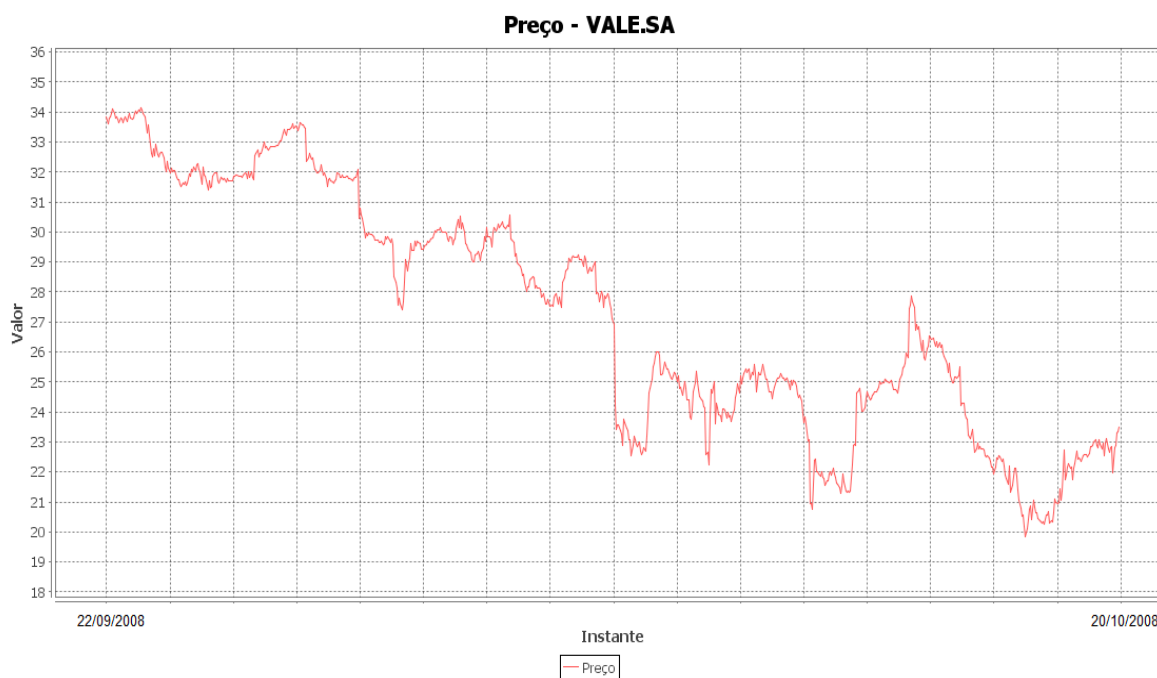
Esta seção mostra os resultados dos agentes obtidos para as ações da VALE5.

### 5.2.2.1 Período de teste número 1

As características do período de teste número 1 é descrito na Tabela 19. O gráfico da variação de preços para este período é representado na Figura 12.

**Tabela 19 - Descrição do período de teste número 1.**

<i>Ação</i>	<i>Início do período</i>	<i>Fim do período</i>	<i>Preço de abertura</i>	<i>Preço de fechamento</i>	<i>Varição</i>
VALE5.SA	22/09/2008	13/10/2008	33.81	24.44	-9.23

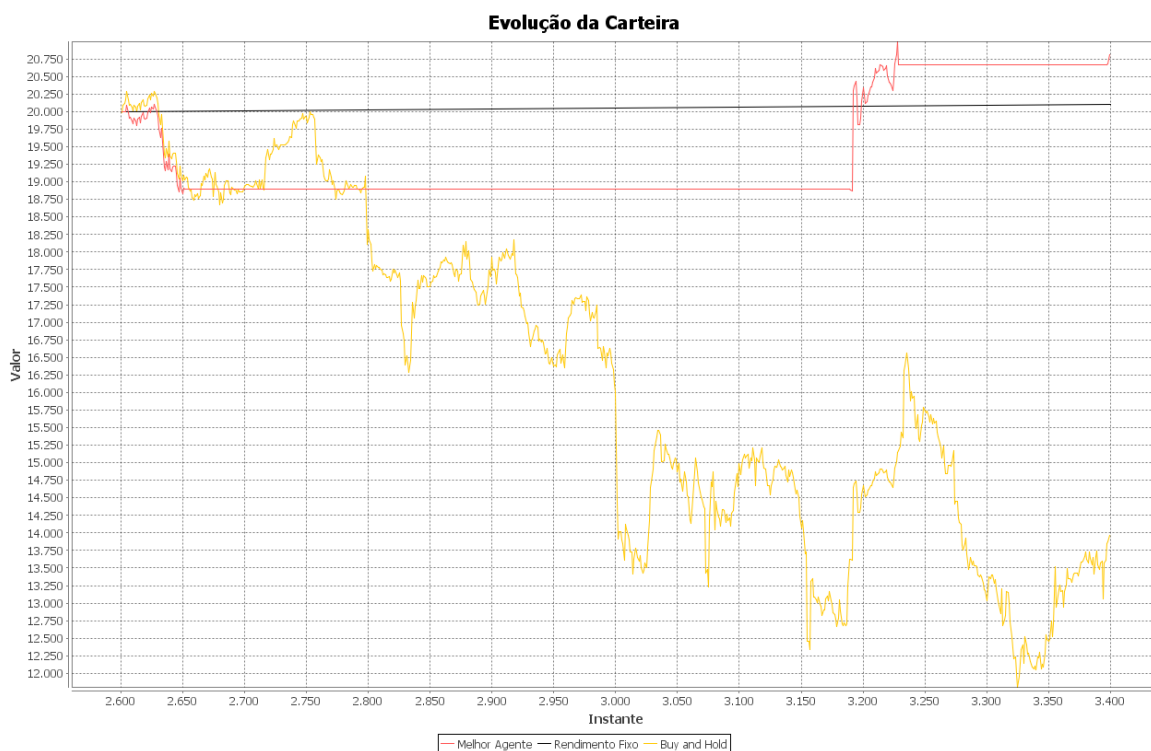


**Figura 12 – Gráfico do preço do período de teste número 1.**

Na Tabela 20 é mostrado o desempenho do melhor agente encontrado para o período número 1. Também está incluído o desempenho do agente de *buy and hold* e do investimento em banco. A Figura 13 representa graficamente este período.

**Tabela 20 - Desempenho dos agentes do período 1.**

<i>Desempenho</i>	<i>Valor Final</i>	<i>Drawdown máximo</i>	<i>Quantidade de Operações</i>	<i>Lucro</i>
Agente Treinado	20823,00	6.31%	5	823,00
Investimento em banco	20100,00	0	0	100,00
<i>Buy and hold</i>	13885,00	33,89%	2	-6115,00



**Figura 13 - Gráfico de desempenho do período 1.**

### 5.2.2.2 Período de testes número 2

As características do período de teste número 2 é descrito na Tabela 21. O gráfico para este período é representado na Figura 14.

Tabela 21 - Descrição do período de teste número 2.

Ação	Início do período	Fim do período	Preço de abertura	Preço de fechamento	Variação
VALE5.SA	14/11/2008	15/12/2008	23.04	24.45	1.41

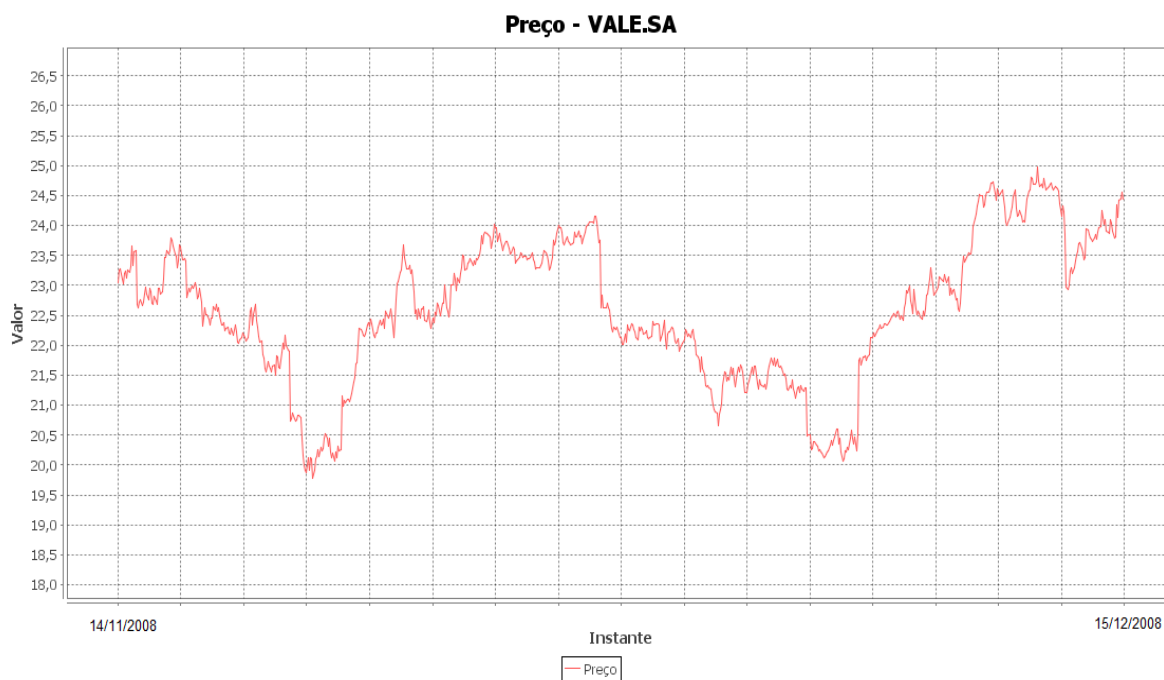


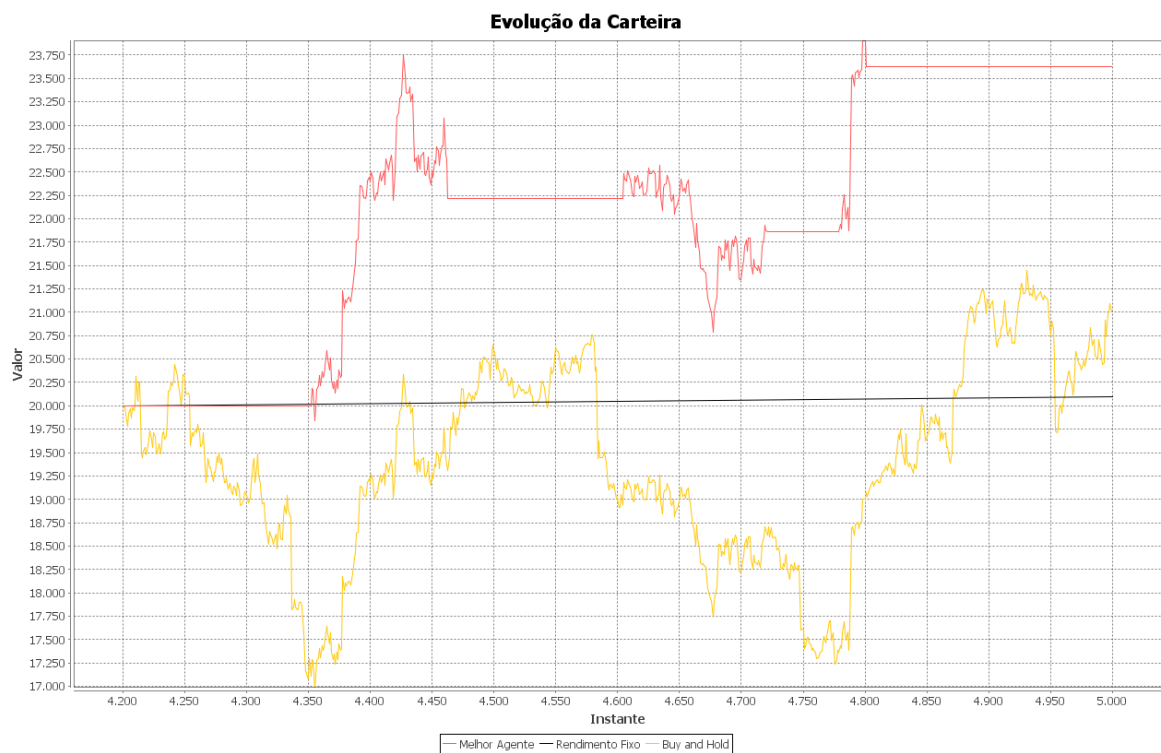
Figura 14 - Gráfico do preço do período de teste número 2.

Na Tabela 22 é mostrado o desempenho do melhor agente encontrado para o período número 2. Também está incluído o desempenho do agente de *buy and hold* e do investimento em banco. A Figura 15 representa graficamente este período.



**Tabela 22 - Desempenho dos agentes do período 2.**

<i>Desempenho</i>	<i>Valor Final</i>	<i>Drawdown</i>	<i>Quantidade de Operações</i>	<i>de Lucro</i>
Agente Treinado	23629,00	12,44%	4	3629,00
Investimento em banco	20100,00	0	0	100,00
Buy and hold	20915,00	15,66%	2	915,00



**Figura 15 - Gráfico de desempenho do período 2.**

### 5.2.2.3 Período de testes número 3

As características do período de teste número 3 é descrito na Tabela 23. O gráfico para este período é representado na Figura 16.

Tabela 23 - Descrição do período de teste número 3.

Ação	Início do período	Fim do período	Preço de abertura	Preço de fechamento	Varição
VALE5.SA	20/10/2008	14/11/2008	23.48	22.96	-0.52

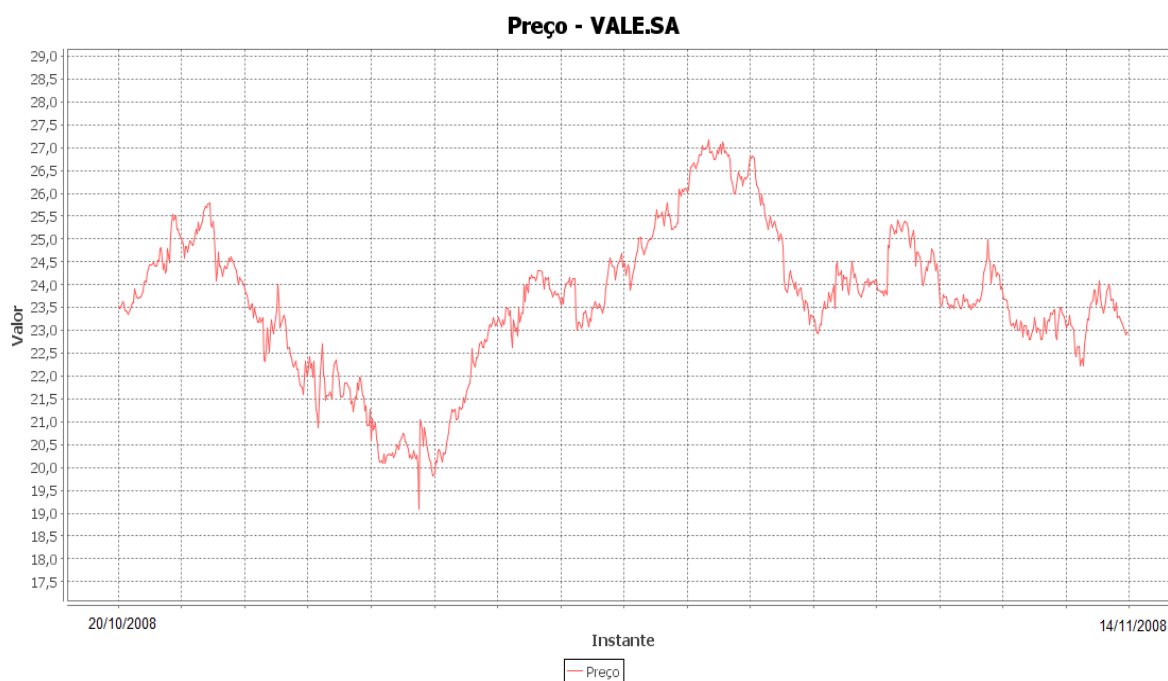
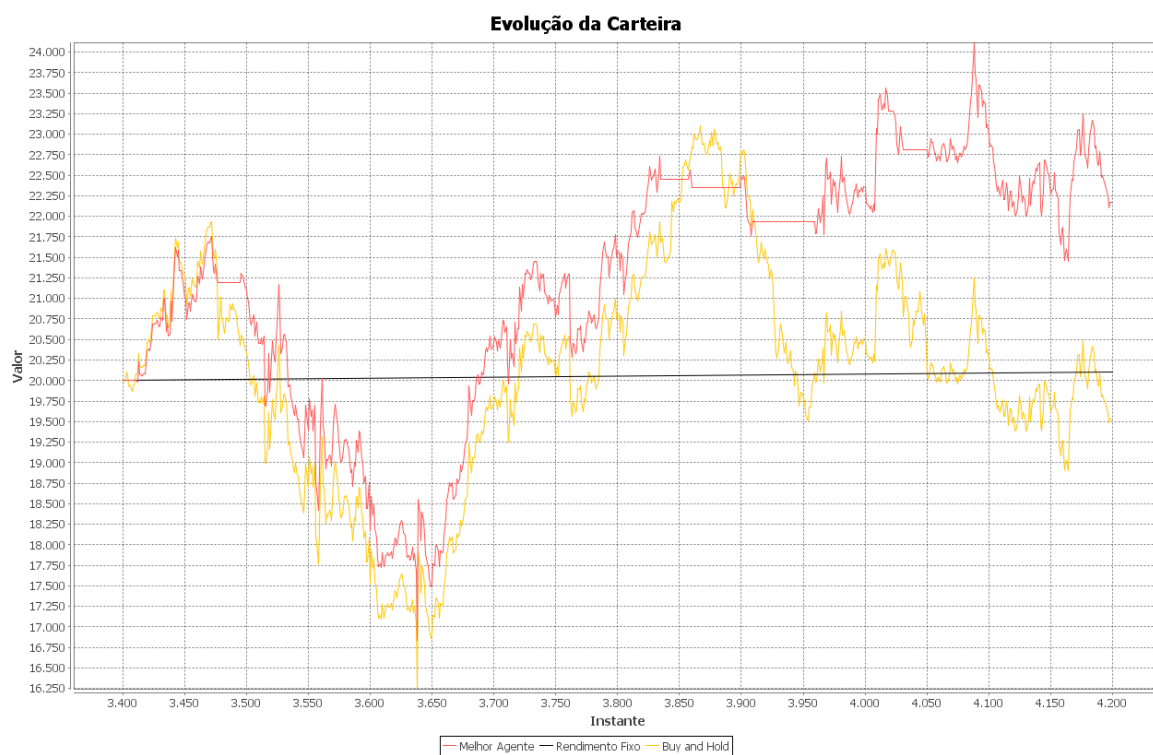


Figura 16 - Gráfico do preço do período de teste número 3.

Na Tabela 24 é mostrado o desempenho do melhor agente encontrado para o período número 3. Também está incluído o desempenho do agente de *buy and hold* e do investimento em banco. A Figura 17 representa graficamente este período.

**Tabela 24 - Desempenho dos agentes do período 3.**

<i>Desempenho</i>	<i>Valor Final</i>	<i>Drawdown máximo</i>	<i>Quantidade de Operações</i>	<i>Lucro</i>
Agente Treinado	22.454,00	22,54%	17	2.454,00
Investimento em banco	20.100,00	0	0	100,00
Buy and hold	19567,00	25,93%	2	-433,00



**Figura 17 - Gráfico de desempenho do período 3.**

## 5.3 Desempenho geral

O desempenho agrupado dos agentes sobre todo período de treinamento disponível é demonstrado na Tabela 25. Os dados da tabela em questão foram gerados através da metodologia descrita na Seção 4.6.

**Tabela 25 - Desempenho geral dos agentes**

<i>Ativo</i>	<i>Início</i>	<i>Fim</i>	<i>Porcentagem de agentes lucrativos encontrados</i>	
PETR4	19/09/2008	27/10/2009	32,57%	
VALE5	22/09/2008	25/03/2010	40,63%	
BBDC	25/09/2008	16/04/2010	31,68%	

<i>Ativo</i>	<i>Média de rendimento mensal</i>	<i>Lucro no período</i>	<i>Lucro buy-and-hold</i>	<i>Lucro investimento bancário</i>
PETR4	2,72%	7.618,00	3.384,00	1.339,72
VALE5	3,32%	12.615,00	6.929,00	1.878,57
BBDC	2,72%	10.320,00	4.595,00	1.987,97

Para cada mês foram executadas 50 instâncias do *hill-climbing* e a porcentagem de acerto é definida como a razão entre o número de agentes que obtiveram lucro no período de teste pelo número de instâncias executadas.

O lucro no período refere-se ao somatório da média do valor obtido de cada agente que obteve lucro no período de teste. A média de rendimento mensal refere-se à razão do lucro no período pelo número de meses usados do ativo em questão.

# 6

## Conclusão

### 6.1 Comparações

Este trabalho investigou a possibilidade de se encontrar regras técnicas lucrativas no mercado acionário da BOVESPA, utilizando a estrutura de árvores de decisões binárias, técnica de Programação Genética e o método *Hill-climbing*. A Programação Genética e o método *Hill-climbing* foram utilizados para encontrar agentes financeiros artificiais lucrativos para as ações PETR4.SA, BBDC.SA e VALE5.SA.

Embora tenha sido utilizada inicialmente a Programação Genética na busca de soluções para o problema, esta se mostrou ineficiente, para a grande quantidade de dados trabalhados. Portanto, optou-se pela aplicação do método *Hill-climbing* e constatou-se que este método obtém resultados satisfatórios em tempo hábil.

De acordo com os resultados demonstrados no Capítulo 5, pode-se corroborar a hipótese  $H_0$  : “É possível encontrar agentes financeiros artificiais que descubram regras técnicas lucrativas mensalmente no mercado acionário da BOVESPA utilizando busca local com o método *hill-climbing* e com estruturas de árvores de decisão”.

Observa-se que agentes treinados segundo a metodologia proposta produzem agentes financeiros que são capazes de obter ganhos significativos quando comparados ao investimento em banco para qualquer período. Em períodos de variação de preço negativa, a estratégia de *buy and hold* sempre causa prejuízo, então o agente treinado também é capaz de obter melhores resultados. Porém em períodos com variação de preço positiva, o agente treinado pode ser menos lucrativo do que a estratégia de *buy and hold*. Ainda assim,

por propiciar um *drawdown* aceitável e uma maior quantidade de operações, o agente financeiro treinado mostra-se confiável.

Contudo, como visto na Tabela 25, ainda são necessárias investigações adicionais sobre o problema, uma vez que o percentual de agentes lucrativos encontrados ainda é baixo.

## 6.2 Discussões e Trabalhos Futuros

Uma questão relevante para os próximos trabalhos nesta área está relacionada à função de avaliação de agentes financeiros descrita aqui, na qual se incluem as funções  $O_b$  e  $O_t$ , relativas, respectivamente, à quantidade de operações boas realizadas e à quantidade de operações totais realizadas pelo agente. Precisa-se investigar a real necessidade destas funções. Primeiramente, porque as operações boas influenciam no lucro final calculado com a função  $L$  e operações ruins podem influenciar na função  $D$ , se forem muito ruins. Além disso, se espera que o agente encontre naturalmente uma quantidade de operações na busca pela solução, que resulte em lucro maior, não sendo necessário que se informe quantas operações o agente deverá realizar. Considerando estes pontos, uma proposta de função *fitness*  $f$  a ser investigada é:

$$(6) \quad f = \text{Lucro}_{\text{liquido}} * (1 - \text{Drawdown}_{\text{maximo}})$$

Outra questão de relevância é o resumo das árvores de decisão contidas no agente financeiro. O agente encontrado no período de treinamento pode conter árvores com profundidade maior que sete níveis, dificultando a compreensão da mesma. Porém é importante notar que nem todos os caminhos destas árvores foram percorridos no período de treinamento, significando que se no período de teste a árvore percorrer este caminho ela tomará uma decisão ainda não vista no treinamento. Isto pode ser um risco, que deve ser evitado. Uma possível solução que precisa ser investigada em trabalhos futuros é a substituição das subárvores não percorridas por folhas contendo valor falso. Este processo

prepara a árvore, resumindo-a, para retornar decisões de espera se no período de teste tentar-se percorrer algum caminho não percorrido no período de treinamento.

Durante o desenvolvimento deste projeto, algumas outras idéias úteis aos trabalhos futuros foram levantadas:

- Permitir agentes financeiros comprarem ações de mais de uma empresa; o que força o agente a não realizar apenas operações absolutas;
- Usar agentes financeiros em conjunto com previsor dos valores máximo e mínimo do dia.

Embora haja interesse nestas duas últimas idéias levantadas, é importante primeiramente obter melhores resultados com a abordagem seguida neste trabalho. Primeiramente, o *software* deveria gerar uma quantidade maior de agentes lucrativos se comparado com os ruins. Para tanto, sugere-se o ajuste de parâmetros do algoritmo, a mudança da técnica de busca no espaço de soluções, a mudança da função de avaliação ou mudança da quantidade de meses para treinamento e teste.





# 7

## Referências Bibliográficas

ACHELIS, S. B. **Technical Analysis from A to Z**. Disponível em: <http://www.equis.com/customer/resources/taaz>>. Acesso em: 6 de dez. 2010.

ANTONIOU, A.; LU, W. **Practical Optimization: Algorithms and Engineering Applications**. Springer, 2007. 669 p.

BMF&BOVESPA. **Mercado de Capitais: Introdução**. São Paulo. 2008. Disponível em: <<http://www.bmfbovespa.com.br/Pdf/merccap.pdf>>. Acesso em: 9 de mai. 2010.

BMF&BOVESPA. **Custos para o Mercado de Ações**. Disponível em: <<http://www.bmfbovespa.com.br/pt-br/regulacao/custos-e-tributos/custos-operacionais/acoes.aspx?Idioma=pt-br>>. Acesso em: 02 jun. 2011.

EIBEN A. E.; SMITH, J. E. **Introduction to Evolutionary Computing**. Natural Computing Series, MIT Press. Berlin: Springer, 2003.

ELDER, A. **Come into My Trading Room: A Complete Guide To Trading**. New York: John Wiley & Sons Inc, 2002. 313 p.

FONTANILLS G.A., GENTILE T. **The Stock Market Course**. New York: John Wiley & Sons Inc, 2001. 464 p.

GUIAINVEST. Disponível em: <<http://www.guiainvest.com.br/grafico>>. Acesso em: 01 dez. 2010.

KAMPOURIDIS M.; CHEN S.; TSANG E. **Market Microstructure: Can Dinosaurs Return? A Self-Organizing Map Approach under an Evolutionary Framework**.

In: Applications of Evolutionary Computation. Berlin: Springer-Verlag , 2011. p. 91-100.

KOZA, R. J. **Genetic programming: On the Programming of Computers by Means of Natural Selection.**1.ed. MIT Press: Cambridge, 1992.

LEE, S.; MOON, B. **Finding Attractive Rules in Stock Markets using a Modular Genetic Programming.** In: GENETIC AND EVOLUTIONARY COMPUTATION CONFERENCE (GECCO 2009), 11, 2009, Montreal. **Anais...** Nova Iorque: ACM, 2009. p. 1933-1934. Disponível em: <<http://portal.acm.org/citation.cfm?id=1570244>>.

PIAZZA, M. C. **Bem Vindo à Bolsa de Valores.** 8. ed. rev. e ampl. São Paulo: Novo Conceito, 2009.

POTVIN J.; SORIANO P.; VALLÉE M. **Generating trading rules on the stock markets with genetic programming.** *Jornaul Computers and Operations Research.* Elsevier Science, Oxford, v.31, n.7, p. 1033-1047, jun. 2004. Disponível em: <<http://portal.acm.org/citation.cfm?id=980038>>

QUINLAN, J.R. **Learning Decision Tree Classifiers.** In *Journal ACM Computing Surveys, New York, NY, USA*, p. 71-72, Mar 1996.

REZENDE, S. O. **Sistemas Inteligentes: Fundamentos e Aplicações.** Editora Manole Ltda., 2005.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial.** Tradução: Vanderberg D. de Souza. 2. ed. Rio de Janeiro: Elsevier, 2004.

SCHULENBURG, S.; ROSS, P. **An Evolutionary Approach to Modeling the Behaviours of Financial Traders.** In: GENETIC AND EVOLUTIONARY COMPUTATION CONFERENCE, 1, 1999. Orlando, Florida. Proceedings of the Genetic and Evolutionary Computation Conference. San Francisco, California: Morgan Kaufmann, 1999. p. 245-253

\_\_\_\_\_. **Strength and Money**: an LCS Approach to Increasing Returns. In: THIRD INTERNATIONAL WORKSHOP, 3, 1996. Paris. Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag, 2001. p. 114-137

\_\_\_\_\_. **A Learning Evolutionary Trading System – LETS**. In: GENETIC AND EVOLUTIONARY COMPUTATION CONFERENCE, 3, 2002. New York. GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference. Morgan Kaufmann, 2002. p. 45-53.

SORNETTE, D. **Why Stock Markets Crash**: Critical Events in Complex Financial Systems. Princeton University Press, 2004. 448 p.

TERMINAL ENFOQUE. Disponível em: <<http://www.enfoque.com.br/>>. Acesso em: 01 dez. 2010.

VOUDOURIS C.; TSANG R. Guided Local Search Joins the Elite in Discrete Optimization. In: FREUDER E. C.; WALLACE R. J. **Constraint Programming and Large Scale Discrete Optimization**. USA: The American Mathematical Society, 2001. p. 29-41.

YAHOO FINANCE. Disponível em: <<http://finance.yahoo.com/>>. Acessado em: 01 jun. 2011