

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

FABIO ALBERTO PROCHNOW

**Programação Genética para Predição de
Séries Temporais Aplicados a Mercados
Financeiros**

Trabalho de Graduação apresentado como
requisito parcial para a obtenção do grau de
Bacharel em Ciência da Computação.

Prof. Dr. Dante Augusto Couto Barone
Orientador

Prof. Msc. Luiz Otávio Vilas Boas Oliveira
Co-orientador

Porto Alegre, janeiro de 2013

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitora de Graduação: Profa. Valquiria Link Bassani

Diretor do Instituto de Informática: Prof. Luíz da Cunha Lamb

Coordenador da Ciência da Computação: Prof. Raul Fernando Weber

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

Primeiramente agradeço aos meus pais, Marcos Prochnow e Mirtha Azambuja, que sempre tentaram me preparar para as oportunidades que a vida hoje me proporciona. Agradeço, também, à minha irmã, Marília Prochnow, e à minha namorada, Luciana Kray, pelo carinho, compreensão e apoio durante os momentos mais críticos deste trabalho.

Agradeço ao professor Dr. Dante Augusto Couto Barone e ao Msc. Luiz Otávio Vilas Boas Oliveira pela orientação, oportunidade e ideias que se desenvolveram desde nossos primeiros encontros e sem as quais não haveria este trabalho.

Aos colegas de curso e de trabalho que compartilharam comigo alguma parcela deste tempo como aluno do curso de Bacharelado em Ciência da Computação e que ajudaram e continuarão cooperando na construção do conhecimento obtido ao longo destes anos e do aprendizado que ainda está por vir.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS.....	6
LISTA DE FIGURAS	7
LISTA DE TABELAS	9
RESUMO	10
ABSTRACT.....	11
1 INTRODUÇÃO	12
1.1 Séries Temporais	12
1.2 Estado da arte da Computação Evolutiva e da Programação Genética..	12
1.3 Objetivos.....	13
1.4 Estrutura do Texto	13
2 OS MERCADOS FINANCEIROS	15
2.1 Mercado Forex.....	15
2.2 Mercado de Ações.....	17
2.2.1 A BM&FBOVESPA	17
2.3 A Análise Técnica	19
3 PROGRAMAÇÃO GENÉTICA	22
3.1 A Origem	22
3.2 Conceitos	22
3.2.1 População e Indivíduo	24
3.2.2 Avaliação dos indivíduos	25
3.3 Seleção e Operadores.....	25
3.3.1 Seleção.....	26
3.3.2 Reprodução.....	26
3.3.3 Cruzamento.....	26
3.3.4 Mutação	27
3.4 Critérios de Parada	28
4 MÉTODOS PARA PREVISÃO DE SÉRIES TEMPORAIS	29
4.1 Naïve	29
4.2 ARIMA	31
4.3 Redes Neurais.....	32
4.4 Regressão Simbólica	34
5 PROCEDIMENTO EXPERIMENTAL	36
5.1 Coleta e Organização dos Dados	36
5.2 O Projeto R.....	36
5.3 O ECJ e a Regressão Simbólica	37
5.3.1 Histórico do Experimento utilizando ECJ	37
5.4 Resultados dos Experimentos Realizados	39
5.4.1 Resultados para EURUSD.....	39

5.4.2	Resultados para GBPUSD	42
5.4.3	Resultados para PETR4	45
5.4.4	Resultados para VALE5	48
5.5	Comparação e Análise dos Resultados	52
6	CONCLUSÕES E SUGESTÕES	54
6.1	Sugestões para Trabalhos Futuros.....	54
	REFERÊNCIAS.....	56
	GLOSSÁRIO.....	61

LISTA DE ABREVIATURAS E SIGLAS

EURUSD	Par Euro – Dólar Americano
GBPUSD	Par Libra Inglesa – Dólar Americano
PETR4	Sigla de um dos tipos de ações da Petrobrás
RMSE	<i>Root Mean Square Error</i> (Raiz quadrada do Erro Médio Quadrático)
VALE5	Sigla de um dos tipos de ações da Companhia Vale do Rio Doce
CVM	Comissão de Valores Mobiliários

LISTA DE FIGURAS

Figura 2.1: Variação do EURUSD entre 01/2005 e 10/2012.	16
Figura 2.2: Ferramenta para divulgação de índices e outras informações.	18
Figura 2.3: O preço desconta tudo.	20
Figura 2.4: O preço segue tendências.	20
Figura 2.5: A história se repete.	21
Figura 3.1: representação de uma árvore.	23
Figura 3.2: fluxograma representando a execução do algoritmo de programação genética.	24
Figura 3.3: Operação de cruzamento mostrando pais e filhos gerados.	27
Figura 3.4: Operação de mutação com estado da árvore antes e depois.	28
Figura 4.1: Gráfico resultante da previsão de uma série utilizando o método Naïve sem mudanças.	30
Figura 4.2: Gráfico resultante da previsão de uma série utilizando o método Naïve com variação constante.	30
Figura 4.3: Exemplo de previsão utilizando ARIMA (3, 4, 5).	32
Figura 4.4: Rede Neural Típica.	33
Figura 4.5: Fluxo para a geração da previsão do próximo valor de uma série.	33
Figura 4.6: Exemplo de previsão de uma série temporal utilizando Redes Neurais.	34
Figura 5.1: Previsão dos 30 próximos valores da série EURUSD usando Naïve.	40
Figura 5.2: Previsão dos 30 próximos valores da série EURUSD usando ARIMA.	40
Figura 5.3: Previsão dos 30 próximos valores da série EURUSD usando Redes Neurais.	41
Figura 5.4: Previsão dos 30 próximos valores da série EURUSD usando Regressão Simbólica.	41
Figura 5.5: Previsão dos 30 próximos valores da série GBPUSD usando Naïve.	43
Figura 5.6: Previsão dos 30 próximos valores da série GPBUSD usando ARIMA.	43
Figura 5.7: Previsão dos 30 próximos valores da série GPBUSD usando Redes Neurais.	44

Figura 5.8: Previsão dos 30 próximos valores da série GBPUSD usando Regressão Simbólica.....	44
Figura 5.9: Previsão dos 30 próximos valores da série PETR4 usando Naïve.....	46
Figura 5.10: Previsão dos 30 próximos valores da série PETR4 usando ARIMA. .	46
Figura 5.11: Previsão dos 30 próximos valores da série PETR4 usando Redes Neurais.....	47
Figura 5.12: Previsão dos 30 próximos valores da série PETR4 usando Regressão Simbólica.....	47
Figura 5.13: Previsão dos 30 próximos valores da série VALE5 usando Naïve.	49
Figura 5.14: Previsão dos 30 próximos valores da série VALE5 usando ARIMA. .	49
Figura 5.15: Previsão dos 30 próximos valores da série VALE5 usando Redes Neurais.....	50
Figura 5.16: Previsão dos 30 próximos valores da série VALE5 usando Regressão Simbólica.....	50

LISTA DE TABELAS

Tabela 2.1: Horários de abertura e fechamento de algumas sessões	16
Tabela 5.1: Valores reais, valores previstos e erros nas previsões para a série EURUSD	42
Tabela 5.2: Valores reais, valores previstos e erros nas previsões para a série GBPUSD	45
Tabela 5.3: Valores reais, valores previstos e erros nas previsões para a série PETR4	48
Tabela 5.4: Valores reais, valores previstos e erros nas previsões para a série VALE5	51
Tabela 5.5: RMSE resultantes de todas as simulações	53

RESUMO

As Séries Temporais podem ser percebidas em diversas formas na natureza e até mesmo nos processos industriais. Nos Mercados Financeiros, por exemplo, pode-se ver nitidamente a formação destas séries. Tanto para os investidores do Mercado Forex quanto para os do Mercado de Ações, o desafio é prever as variações destas séries e obter o maior lucro possível destes comportamentos. Para isso, foi criada a Análise Técnica, que consiste de fundamentos e ferramentas de análise gráfica para auxiliar os investidores na hora de tomar uma decisão.

Ao encontro disso, surgem os métodos clássicos de predição de Séries Temporais como o Naïve, o ARIMA e, nos últimos tempos, as próprias Redes Neurais. Por outro lado, a Programação Genética vem se destacando em inúmeras aplicações práticas e, dentre as possibilidades de uso desta, está a Regressão Simbólica. Por esse motivo, realizaram-se experimentos comparativos entre os métodos mais utilizados para a previsão destas séries e a própria predição por Regressão Simbólica. Para isso, foram coletadas séries referentes aos artigos mais movimentados nos Mercados de Ações e Forex como as ações PETR4 e VALE5 e os pares EURUSD e GBPUSD.

Por fim, percebe-se que a Regressão Simbólica pode ser mais um aliado dos investidores na busca pelo lucro e, embora não haja um método perfeito entre os testados, todos eles devem ser utilizados em conjunto de forma a reduzir os riscos na hora de investir.

Palavras-chave: Programação Genética, Séries Temporais, Previsão, Análise Técnica, Mercados Financeiros.

Genetic Programming for Time Series Forecasting Applied to Financial Markets

ABSTRACT

The Time Series can be seen in different ways in the nature or even in the industrial processes. In the Financial Markets, for example, it can be clearly seen the formation of these series. For the investor of the Forex Market as for the investors of the Stock Market the challenge is to forecast the variations of these series in order to acquire the best profit rate from these behaviors. The Technical Analysis was created with this objective and it provides principles and graphical analysis tools to help the investors on their decisions.

In this sense, some classical methods to forecast Time Series such as the Naïve method, the ARIMA method and even the Neural Networks are widely used. On the other hand, the Genetic Programming has been very successfully applied to many practical applications and one of its possible uses is the Symbolic Regression. For this reason, some experiments were realized to make a comparison among the classical methods of Time Series forecasting and the prediction using the Symbolic Regression method. In order to do these experiments, some Time Series of the most active articles of the Stock and Forex Markets such as the PETR4 and VALE5 shares and the EURUSD and GBPUSD pairs were gathered.

At the end of the tests, it was seen that the Symbolic Regression can be considered another ally of the investors seeking high profit rates and, although there isn't a perfect method among the ones used in this work, all of them should be used together in order to reduce the risks when it's time to invest.

Keywords: Genetic Programming, Time Series, Forecasting, Technical Analysis, Financial Markets.

1 INTRODUÇÃO

As sessões a seguir apresentam uma visão geral acerca dos tópicos básicos necessários para a compreensão deste documento bem como os objetivos do trabalho e a forma como esta monografia está organizada para alcançá-los. Para maiores detalhes sobre os assuntos abordados, recomenda-se consultar as referências utilizadas.

1.1 Séries Temporais

Uma série temporal é uma sequência ordenada de valores observados para uma determinada variável medida em intervalos iguais de tempo (NIST/SEMATECH, 2012). O ordenamento da sequência pode se dar em outras dimensões como, por exemplo, o espaço. É importante ressaltar que uma característica intrínseca das séries temporais é que as suas observações de dados são interdependentes e, portanto, manter a ordem em que os dados foram gerados é imprescindível (WEI, 2006).

Alguns dados das séries temporais podem ser obtidos, por exemplo, através de observações de processos industriais ou até mesmo de métricas corporativas (NIST/SEMATECH, 2012). Uma forma de obter os dados para estas séries é através da observação e análise de alguns eventos, tais como sinais elétricos na engenharia e a velocidade dos ventos e a temperatura diária na meteorologia (WEI, 2006).

A análise destas séries temporais apresenta inúmeras aplicações como a previsão econômica, a análise do mercado de ações, o controle de processos e de qualidade, entre outras (NIST/SEMATECH, 2012). Dessa forma, existem vários objetivos no estudo de séries temporais e, dentre eles, estão o entendimento e descrição do mecanismo de geração destes valores, a previsão de futuros valores e a otimização do controle de um sistema (WEI, 2006). Em especial, a previsão de valores futuros e o estudo comparativo de um método específico para esse processo utilizando programação genética são o foco deste trabalho.

1.2 Estado da arte da Computação Evolutiva e da Programação Genética

As origens da Computação Evolutiva estão na década de 1950, em artigos resultantes dos trabalhos de Bremermann, Friedberg e Box. Todavia, foi apenas na década de 1970 com os trabalhos de Holland, Rechenberg, Schwefel e Fogel que a Computação Evolutiva começou a se tornar mais popular na comunidade científica (BÄCK; SCHWEFEL, 2007). Nos dias atuais se observa um crescimento exponencial no número de publicações em torno do assunto, demonstrando a importância não somente científica, mas também econômica do tópico (BÄCK; SCHWEFEL, 2007).

A ideia da Programação Genética em especial, foi explorada por John Koza, em (KOZA, 1992), baseado na ideia de Algoritmos Genéticos de John Holland. O objetivo da Programação Genética é utilizar os conceitos de evolução de Darwin como uma forma de induzir um programa a encontrar uma boa solução para problemas específicos (ZHANG; CHENG; YUAN, 2009). Maiores detalhes sobre a Programação Genética e seus procedimentos serão abordados em capítulos posteriores.

O estado atual da Computação Evolutiva pode ser caracterizado por uma série de itens contendo desde o reconhecimento do seu potencial de resolver problemas até a falta de uma definição mais formal sobre o domínio e os limites do paradigma. No que diz respeito às pesquisas nesta área, nota-se uma inclinação dos pesquisadores à busca de diferentes aplicações para esse paradigma. Ainda, percebe-se que as aplicações da Computação Evolutiva em diferentes áreas têm produzido resultados surpreendentemente bons de forma que, algumas vezes, chegam a ser superiores aos resultados encontrados em técnicas mais clássicas da área aplicada (BÄCK; SCHWEFEL, 2007).

Por fim, os estudos mais atuais envolvendo a Programação Genética envolvem diversas áreas como aplicações financeiras (LI, 2000), programas envolvendo lentes ópticas (JONES; AL-SHAKRAN; KOZA, 2006), aplicações em Geografia (SPECTOR; KLEIN, 2006), evolução de redes neurais (BANZHAF; LEIER, 2006), reconhecimento de formas de objetos para a robótica (TUCI; MASSERA; NOLFI, 2010), modelagem de sistemas de agentes cooperativos em jogos (SIM; WANG, 2005). Para estudar mais sobre a Programação Genética ou a própria Computação Evolutiva, recomenda-se a leitura de outras obras dos autores citados na sessão como (KOZA, 1992) e (BÄCK; SCHWEFEL, 2007), além dos artigos da série *Genetic Programming Theory and Practice*, da editora Springer.

1.3 Objetivos

Aplicando os conceitos de Séries Temporais em Mercados Financeiros, este trabalho utiliza Programação Genética para modelar o comportamento dos valores de alguns produtos como, por exemplo, ações e outros ativos. A partir do modelo gerado, espera-se prever com razoável precisão os próximos valores para as séries consideradas.

Com o objetivo de compreender a qualidade da solução, este trabalho realizará testes com outras formas de aproximação de séries ou pontos, reutilizando os mesmos dados empregados para os experimentos com Programação Genética. Por fim, os resultados serão comparados para definir as diferenças dos erros de cada uma das previsões.

1.4 Estrutura do Texto

A presente monografia está estruturada em sete capítulos, incluindo a introdução. Os outros seis capítulos são descritos a seguir:

O capítulo dois apresenta os conceitos de mercado financeiro e duas das principais fontes de séries temporais dentro deste mercado, o mercado de ações e o mercado cambial. Ainda neste capítulo, é apresentada uma das maneiras mais utilizadas para tentar prever o comportamento deste mercado, a Análise Técnica.

O terceiro capítulo se aprofunda nos princípios da Programação Genética, iniciando pelos fundamentos do método e passando pelos conceitos de populações, gerações e operadores.

O quarto capítulo consta de outras técnicas utilizadas para aproximar pontos de uma função ou série. Dentre os tópicos abordados neste capítulo estão o método Naïve, o método ARIMA, a utilização de Redes Neurais e, por fim, a Redução Simbólica, realizada através da Programação Genética.

O capítulo cinco explica todo o procedimento experimental realizado pelo autor. Inicia-se pela forma como os dados foram coletados e organizados, continuando pelos experimentos utilizando outras formas de aproximar as séries coletadas, bem como a descrição da ferramenta utilizada para isso. Em seguida é feita a descrição do experimento utilizando Programação Genética bem como da API ECJ, utilizada para isso. Por fim, é feita uma comparação dos resultados obtidos com os experimentos descritos e a análise destes resultados.

Ao final do documento, nos capítulos seis e sete, são apresentadas as conclusões deste trabalho juntamente com sugestões para trabalhos futuros e as referências utilizadas na realização do mesmo.

2 OS MERCADOS FINANCEIROS

A principal função da economia é garantir que o capital está sendo alocado eficientemente e, para isso, os investimentos devem ser priorizados para áreas onde os resultados sejam melhores. Dessa forma, acredita-se que os Mercados Financeiros são uma forma de garantir a distribuição adequada dos investimentos (WURGLER, 2000).

Os Mercados Financeiros existem há muito tempo, tendo surgido antes mesmo de existir qualquer forma organizada de troca (LEVINSON, 2005). Nos tempos em que a humanidade focava no plantio de milho, por exemplo, um agricultor que tivesse uma colheita ruim, deveria buscar crédito na forma de sementes para o plantio da próxima safra através de outros agricultores. Por outro lado, se este mesmo agricultor tivesse tido uma boa colheita, ele deveria optar entre trocar o produto por outros alimentos ou estocá-los como sementes para a próxima safra. Esse tipo de dúvida está sempre presente na vida de um investidor ou de um *trader* (investidor de curtíssimo prazo) de commodities dos tempos atuais (LEVINSON, 2005).

Costuma-se dizer que os mercados financeiros estão constantemente em um processo de mudanças (DAVIDSON, 2009). Essas mudanças que envolvem contrações e expansões da economia seguem tendências que são ditadas pelas expectativas dos investidores e *traders* acerca do futuro (PRING, 2002).

Ainda que dentre os mercados financeiros existam outros diferentes mercados e que estes operem de formas distintas entre si, todos eles possuem as mesmas funções como, por exemplo, a fixação de preços, levantamento de capital, investimentos, entre outras (LEVINSON, 2005). Nas sessões a seguir veremos dois mercados financeiros em especial: o Mercado Forex, que é atualmente um dos maiores mercados do mundo (CVM, 2012), e o Mercado de Ações, com uma abordagem superficial do mercado brasileiro.

2.1 Mercado Forex

O preço de qualquer produto em qualquer país é expresso em termos de uma determinada moeda. Entretanto, o valor real de uma unidade monetária só pode ser determinado em termos de alguma outra referência que, em geral, é tomada na forma de outra moeda. Ao valor determinado por esse referencial se dá o nome de Taxa de Câmbio, ou Taxa Cambial (LEVINSON, 2005). A Figura 2.1 mostra um gráfico da variação Taxa Cambial entre o Euro e Dólar Americano (EURUSD) entre janeiro de 2005 e outubro de 2012.



Figura 2.1: Variação do EURUSD entre 01/2005 e 10/2012 (ZULUTRADE, 2012).

O mercado onde as moedas são comercializadas pode ser chamado de vários nomes como Mercado Cambial ou Mercado Forex (abreviatura para *Foreign Exchange*, que significa Câmbio Estrangeiro) ou Mercado FX (abreviatura para Forex) (PONZY, 2007). Sendo ele capaz de influenciar muitos pontos específicos da economia global, como taxas de inflação e o fluxo de investimentos internacionais, o Mercado Forex é considerado o maior mercado do mundo (LEVINSON, 2005), onde se estima que o volume de dinheiro comercializado atinja 1,9 trilhões de Dólares ao dia (LEVINSON, 2005; PONSI, 2007). As moedas mais comercializadas atualmente são o Dólar Americano, a Libra Esterlina e o Euro. Depois dessas, estão o Franco Suíço, o Yen, o Dólar Australiano, o Dólar Neozelandês e a Coroa Norueguesa, seguidos das moedas dos mercados emergentes (DAVIDSON, 2009).

Este mercado não possui uma sede fixa para as transações, podendo estas ocorrer a partir de diversas partes do mundo (DAVIDSON, 2009) e a qualquer momento, pois o Mercado Forex fica aberto 24 horas por dia (PONSI, 2007). Diz-se que ele está sempre aberto devido à existência de diferentes sessões de mercados cujos horários se sobrepõem. As três principais sessões são a Asiática, a Europeia e a Americana. Outros países ainda abrem suas sessões em períodos diferentes, como é o caso da Austrália ou da Nova Zelândia (PONSI, 2007). A Tabela 2.1 mostra os horários de abertura e fechamento de cada uma das quatro sessões citadas anteriormente, considerando o fuso horário do meridiano de Greenwich:

Tabela 2.1: Horários de abertura e fechamento de algumas sessões

Mercado Forex	Inicia Atividades	Termina atividades
Austrália / Nova Zelândia	2100 GMT	500 GMT
Japão / Ásia	2300 GMT	700 GMT
Europa / Grã-Bretanha	700 GMT	1600 GMT
Estados Unidos / Canadá	1200 GMT	2100 GMT

Fonte: (PONSI, 2007)

Por fim, outra característica interessante e ao mesmo tempo perigosa do Mercado Forex é a possibilidade de alavancagem, que é utilizada quando um investidor deseja aumentar o total investido nos lotes sem necessariamente adicionar mais dinheiro próprio à compra (LEVINSON, 2005). Algumas empresas especializadas no mercado oferecem alavancagens de mais de 200 para 1, isso significa que com apenas US\$ 50,00 seria possível fazer transações de até US\$ 10.000,00 (DAVIDSON, 2009). O uso desta característica permite ao investidor aumentar significativamente o seu lucro, mas, por outro lado, ela também permite um aumento de mesma intensidade nas perdas (PONSI, 2007).

2.2 Mercado de Ações

O Mercado de Ações é um segmento de Mercado de Capitais que por sua vez faz parte do Mercado Financeiro (GODOY; MEDINA; JUNIOR, 2006). No Mercado de Ações são realizadas operações financeiras de prazos que variam entre médio e longo, possibilitando que as empresas de capital aberto financiem seus investimentos (GODOY; MEDINA; JUNIOR, 2006). Como o próprio nome sugere, neste último são negociadas ações as quais, por sua vez, são pequenas partes de uma empresa (DAVIDSON, 2009).

As ações são sempre nominativas e costumam ser divididas em dois tipos principais (GODOY; MEDINA; JUNIOR, 2006; BOVESPA, 2012):

- **ORDINÁRIAS (ON):** ações que dão, ao acionista, o direito ao voto em assembleias deliberativas da empresa.
- **PREFERENCIAIS (PN):** ações que são prioritárias no recebimento de valores provenientes de resultados ou, eventualmente, da liquidação da empresa. Esse tipo de ação geralmente não garante direito a voto em assembleias como uma ação ordinária.

As ações são negociadas em Bolsas de Valores as quais promovem as condições necessárias para o funcionamento de mercado. No que se refere ao Brasil, a principal Bolsa onde ações são negociadas é a Bolsa de Valores de São Paulo, popularmente conhecida por Bovespa (GODOY; MEDINA; JUNIOR, 2006). Para operar na Bovespa, o investidor necessita de uma empresa intermediária, conhecida como corretora, que seja registrada junto à Bolsa. As corretoras devem ser autorizadas e fiscalizadas pelo Banco Central e pela Comissão de Valores Mobiliários (BM&FBOVESPA, 2012a; GODOY; MEDINA; JUNIOR, 2006).

2.2.1 A BM&FBOVESPA

A BM&FBOVESPA surgiu em 2008 com a fusão entre a Bolsa de Valores de São Paulo (Bovespa) e a Bolsa de Mercados & Futuros. Com suas sedes nas cidades de São Paulo, Rio de Janeiro e outros países, é considerada uma das maiores Bolsas do mundo e, hoje, desenvolve ferramentas de forma a dar suporte à negociação de ações e derivados (BM&FBOVESPA, 2012b). A BM&FBOVESPA é a única bolsa de valores, mercadorias e futuros em operação no Brasil e, ainda, é um dos responsáveis por fomentar o mercado de capitais brasileiro (BM&FBOVESPA, 2012b).

Além de prover software para suporte, a BM&FBOVESPA divulga índices relacionados ao desempenho das ações de alguma empresa ou de um grupo de empresas de um setor. Os preços das ações e os índices costumam variar tanto por fatores relacionados às empresas quanto por fatores externos, como o crescimento do país, dos níveis de emprego e da taxa de juros. Dessa forma, o comportamento de algumas das ações de um índice pode divergir do comportamento do próprio índice. (BM&FBOVESPA, 2012c)

A Figura 2.2 mostra a variação do índice IBOVESPA divulgado pela BM&FBOVESPA no dia 9 de novembro de 2012. Percebe-se que a ferramenta ainda divulga outros dados como as ações que tiveram maiores altas e maiores baixas em um determinado período. Essa ferramenta está disponível em (BM&FBOVESPA, 2012d).

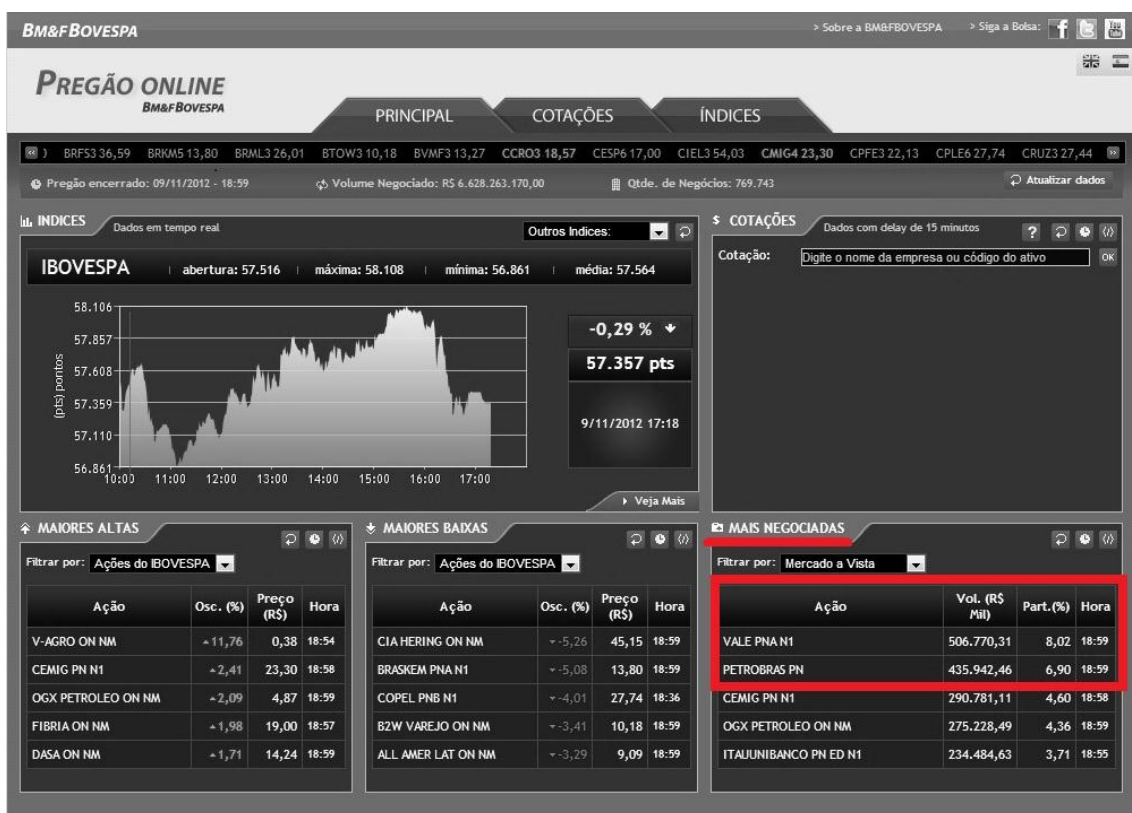


Figura 2.2: Ferramenta para divulgação de índices e outras informações (BM&FBOVESPA, 2012d).

Finalmente, é possível verificar na área destacada no canto inferior direito da imagem que as ações da Companhia Vale do Rio Doce (VALE PNA N1, também conhecida por VALE5) e da Petrobrás (PETROBRAS PN, também conhecida por PETR4) foram as mais negociadas no dia. Os valores históricos das ações destas empresas serão utilizados posteriormente no capítulo 4.

2.3 A Análise Técnica

O principal argumento para a existência da Análise Técnica é o da repetição da história, com possíveis pequenas diferenças, mas com padrões gráficos que se repetem ao longo do tempo (MATSURA, 2007). Sendo assim, a análise técnica estuda o comportamento das ações do mercado se utilizado de vários tipos de gráficos, com propósito de prever as tendências dos preços no futuro (MURPHY, 1999). Como objetivo imediato, a Análise Técnica busca prever o comportamento do preço de um ativo no futuro possibilitando uma decisão quanto à compra ou à venda deste, sempre visando obter o lucro nas transações (ROCKEFELLER, 2011).

A análise técnica se baseia em três fundamentos principais (MATSURA, 2007; MURPHY, 1999):

1. O Preço desconta tudo;
2. O Preço segue tendências;
3. A história se repete.

O primeiro fundamento procura tornar clara a informação de que o preço de um ativo reflete não somente as flutuações da economia de uma empresa ou país, mas também fatores políticos e geográficos que possam ser relevantes aquele ativo (MURPHY, 1999). O segundo fundamento sugere que o preço é o reflexo das atitudes dos investidores (LI, 2000) e que eles tendem a seguir uma tendência principal, ainda que possam oscilar entre alguns valores (MATSURA, 2007). Por fim, o terceiro item tenta explicitar que quando os investidores ou *traders* se deparam com padrões já vistos anteriormente, seu comportamento tende imitar decisões tomadas na ocasião anterior devido à própria natureza estática humana (LI, 2000). As Figuras 2.3, 2.4 e 2.5, respectivamente, exemplificam os fundamentos expostos.



Figura 2.3: O preço desconta tudo (MATSURA, 2007).

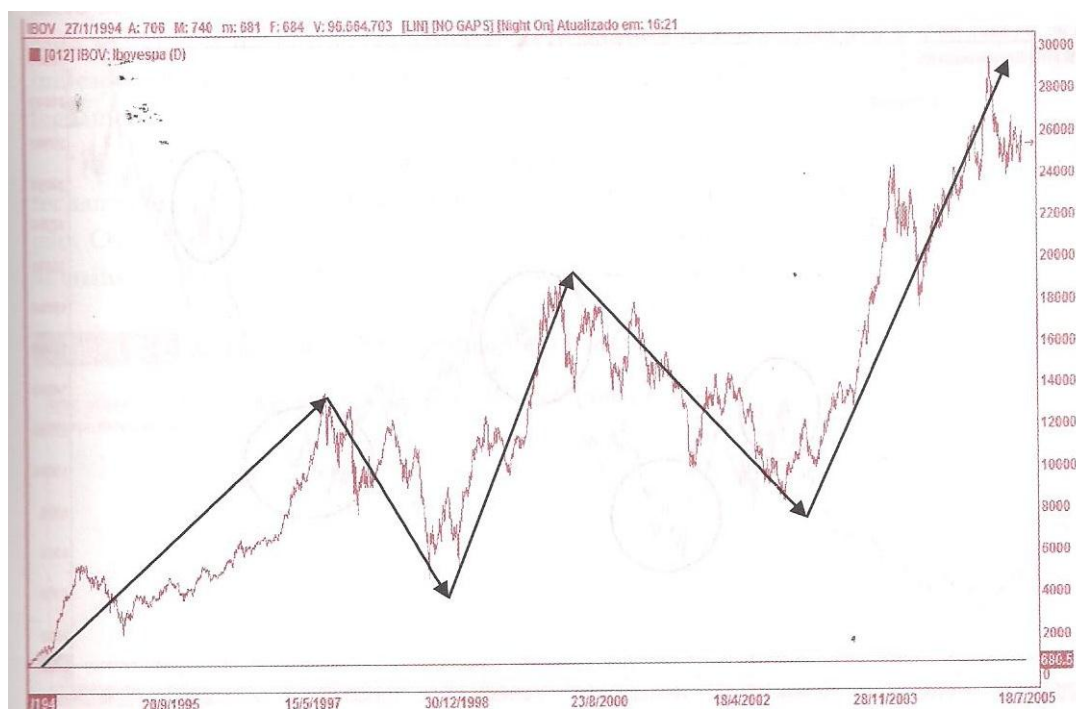


Figura 2.4: O preço segue tendências (MATSURA, 2007).

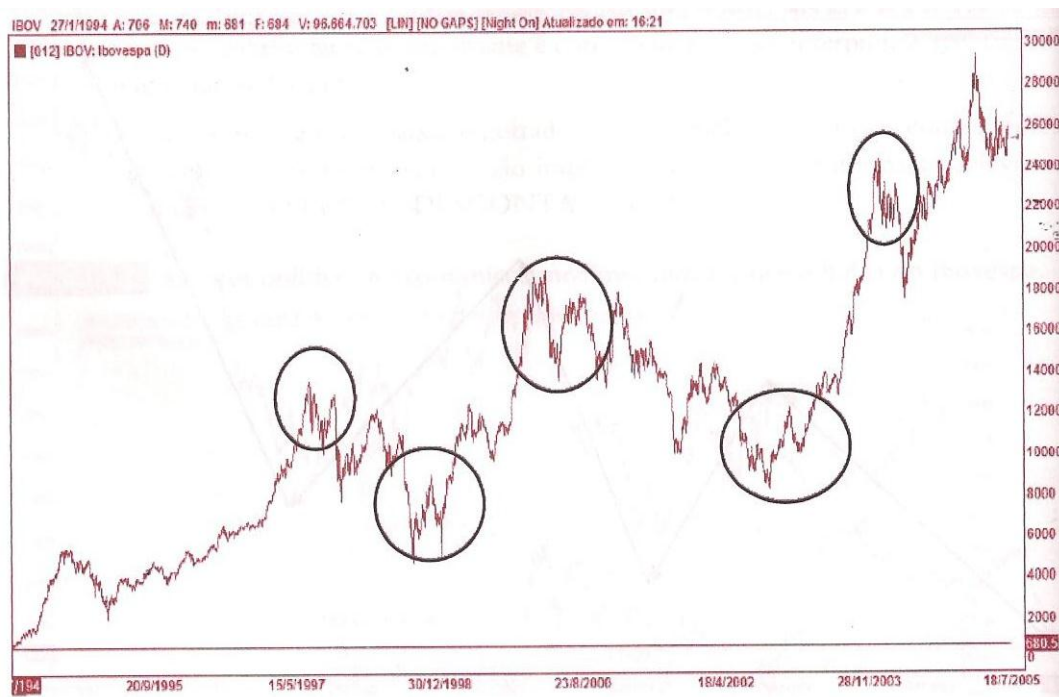


Figura 2.5: A história se repete (MATSURA, 2007).

Para maiores detalhes sobre a Análise Técnica, recomenda-se a leitura das obras completas de (MATSURA, 2007), (MURPHY, 2000) e (ROCKEFELLER, 2011). No capítulo seguinte serão vistos os principais conceitos de Programação Genética e como isso se aplica dentro da Análise Técnica, ou na predição dos valores futuros propriamente ditos.

3 PROGRAMAÇÃO GENÉTICA

Como exposto no capítulo 1, as publicações envolvendo Computação Evolutiva e Programação Genética têm se desenvolvido muito nas últimas décadas, principalmente na área experimental. A Programação Genética foi escolhida como método base para este trabalho justamente pelo fato das aplicações práticas deste método tenderem a apresentar bons resultados.

Neste capítulo, serão apresentados os conceitos básicos necessários para o entendimento da Programação Genética de forma a facilitar o entendimento do trabalho que foi realizado e que será apresentado no capítulo 5.

3.1 A Origem

Como citado anteriormente, a Programação Genética surgiu a partir de uma publicação de (KOZA, 1992), baseando-se nos conceitos de Algoritmos Genéticos. Seu principal objetivo é induzir programas de computadores baseando-se em conceitos existentes na teoria da evolução de Darwin. Em geral, esses conceitos são desenvolvidos na forma de operadores genéticos tais como a seleção, o cruzamento e a reprodução (ZHANG; CHENG; YUAN, 2009), os quais serão vistos na próxima seção.

Na Programação Genética, as estruturas que representam o indivíduo em processo de adaptação são programas hierárquicos de tamanhos e formatos variáveis e o próprio paradigma foi criado de forma a aperfeiçoar a forma de representação dos Algoritmos Genéticos, onde as estruturas são muito menos complexas. Na apresentação do paradigma são sugeridas muitas aplicações como a otimização de controles, planejamento de caminhos, regressão simbólica, programação automatizada, descoberta empírica e previsão, diferenciação e integração simbólica, indução de árvores de decisão, entre outras (KOZA, 1992). Neste trabalho em especial, utilizaremos a Regressão Simbólica para a obtenção da função que aproxima a série temporal e, posteriormente, tentar prever os próximos valores desta série. Mais informações sobre este processo serão vistos na sessão 4.4.

3.2 Conceitos

Diferentemente dos algoritmos genéticos, onde cada gene costuma ser representado por um vetor de valores binários, na Programação Genética cada gene costuma ser modelado por uma árvore que, por sua vez, possui toda a informação relevante sobre um determinado indivíduo (SESHADRI, 2003). A Figura 3.1, a título de exemplo, é uma árvore que representaria a expressão matemática x^2+2x+1 . Semanticamente, esta árvore poderia ser a aproximação de uma função e a sua função de *fitness* (valor que representa

o quão próximo o indivíduo está da solução ideal) poderia ser uma função que representa o erro da aproximação. (SESHADRI, 2003).

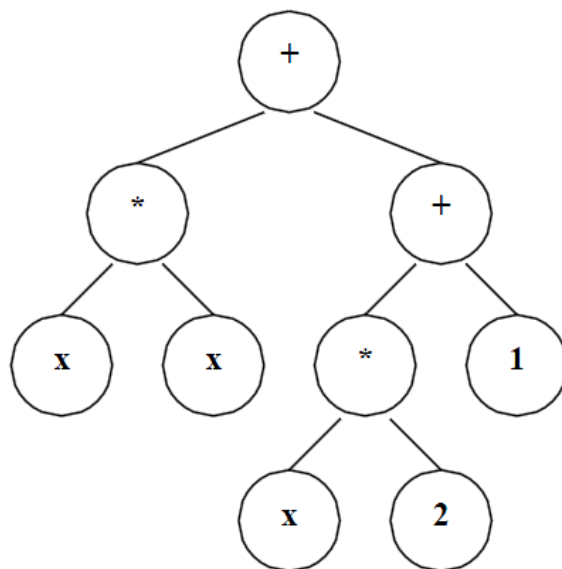


Figura 3.1: representação de uma árvore (SESHADRI, 2003).

A programação genética é uma forma de busca no espaço de soluções, e seu algoritmo poderia ser resumido em 3 passos (KOZA, 1992; LI, 2000):

1. Definir uma população inicial de indivíduos formados por combinações randômicas de funções e terminais.
2. Executar as seguintes tarefas até satisfazer o critério de parada (como o número máximo de gerações, por exemplo):
 - a. Avaliar cada membro da população, atribuindo-lhe um valor de *fitness*.
 - b. Criar uma nova população a partir da existente aplicando, com diferentes probabilidades, operadores de seleção e reprodução apropriados.
3. Escolher o indivíduo que melhor representa o resultado da execução do algoritmo.

Os passos citados acima também podem ser representados em um fluxograma como apresentado na Figura 3.2. Alguns dos conceitos apresentados nesses passos serão vistos nas seções a seguir.

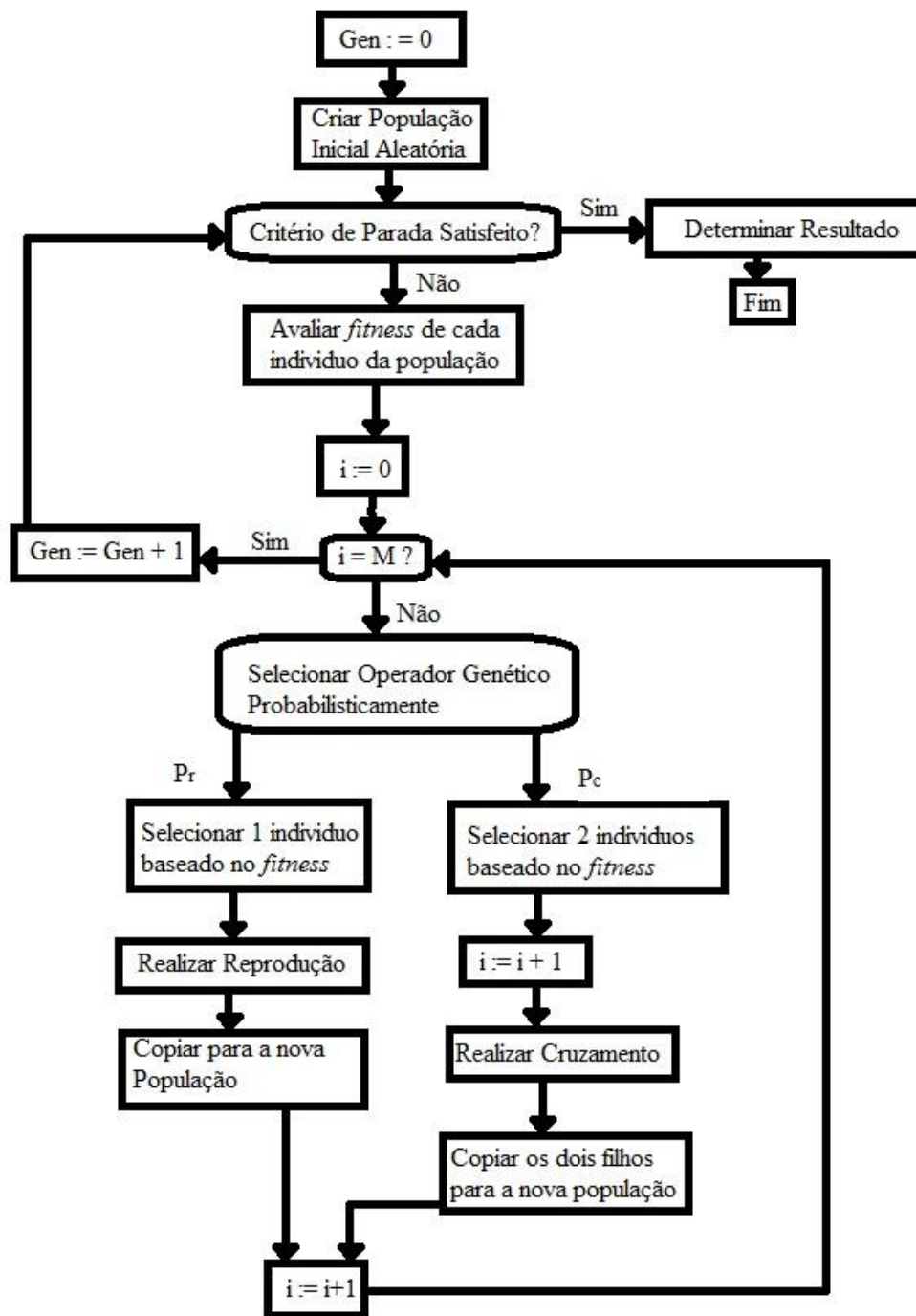


Figura 3.2: fluxograma representando a execução do algoritmo de programação genética (KOZA, 1992).

3.2.1 População e Indivíduo

Uma população é uma coleção de indivíduos gerados aleatoriamente (ZHANG; CHENG; YUAN, 2009) ou, ainda, as estruturas que são adaptadas ao longo do tempo e que representam pontos individuais do espaço de busca (KOZA, 1992). Para a formação de cada indivíduo, deve-se definir o conjunto de funções possíveis (como os operadores aritméticos, funções matemáticas, operações booleanas, funções de iteração, entre outras) e o conjunto de terminais (que inclui as constantes e as variáveis) (KOZA, 1992).

Existem dois métodos básicos para a geração de indivíduos: o método *full* e o método *grow*. Para ambos se define como parâmetro um tamanho máximo da árvore a ser gerada, mas a grande diferença entre estas técnicas é que a primeira gera árvores de forma que a distância entre a raiz e qualquer uma das folhas é a mesma (LI, 2000) e este valor é sempre igual ao parâmetro informado, enquanto a segunda não requer que este critério seja atingido, bastando que os caminhos tenham tamanho menor ou igual ao valor do parâmetro (LI, 2000). Baseado nessas duas técnicas surge uma terceira, a *ramped half-and-half* (KOZA, 1992).

O método generativo *ramped half-and-half* consiste da criação de uma distribuição homogênea de árvores de tamanhos que variam entre 2 e a altura máxima especificada. Por exemplo, considerando uma altura máxima de valor 6, para cada altura entre 2 e 6, teríamos um número de árvores equivalente a 20% do total da população. Adicionalmente, para cada valor altura considerado, metade das árvores são geradas com o método *full* e a outra metade são produzidas via método *grow*. (KOZA, 1992)

A avaliação de cada indivíduo da população é dada a partir de seu valor de *fitness*, que será abordado a seguir.

3.2.2 Avaliação dos indivíduos

Um dos principais passos no uso da Programação Genética é identificar e implementar uma forma de avaliar o quão bom indivíduo é em resolver um determinado problema (KOZA, 1992). Dessa forma, a função de *fitness* é diretamente responsável pelo desempenho do sistema uma vez que pode ser bastante custoso fazer cálculos baseados em uma árvore (LI, 2000). É bastante comum que a função de *fitness* seja representada por uma função única, baseada em apenas um critério (LI, 2000). Para os casos de redução simbólica, que serão vistos nos próximos capítulos, a forma mais comum de avaliar o *fitness* de um indivíduo é através do cálculo do erro gerado ao aproximar uma determinada série (KOZA, 1992).

Com a finalidade de comparar os resultados obtidos e escolher uma função de *fitness* mais apropriada, este trabalho utiliza o RMSE (*Root Mean Square Error*) considerando o fato de este método ser o mais aceito (HOANG et al., 2006). A seguir serão abordadas algumas operações aplicadas às populações de modo a encontrar ou gerar outros indivíduos que apresentem uma melhor função de *fitness*.

3.3 Seleção e Operadores

Existem basicamente dois operadores primários para modificar os indivíduos de uma determinada população e cerca de cinco operadores secundários que podem ser utilizados ocasionalmente (KOZA, 1992). Na Programação Genética os operadores são muito similares aos utilizados nos Algoritmos Genéticos, diferenciando-se apenas na adaptação do seu funcionamento para operar em árvores. A reprodução e o *crossover* são considerados os principais operadores genéticos, entretanto a mutação também pode ser utilizada, guardado os devidos cuidados (LI, 2000).

Para a aplicação destes operadores, é necessário selecionar os indivíduos da população que irão sofrer alteração naquela geração. Dessa forma, existem métodos para escolher os candidatos que possam levar a melhores resultados (KOZA, 1992). Será visto, a seguir, como pode ser feita a seleção de um indivíduo e como funciona cada um destes operadores citados.

3.3.1 Seleção

Existem vários métodos para selecionar um indivíduo de uma população. Dentre eles, estão o método da **posição** (*Rank Method*) e o método do **torneio** (*Tournament Method*) (KOZA, 1992).

O método da posição nada mais é que uma classificação dos valores de *fitness* dos indivíduos da população, sem necessariamente considerar o seu valor numérico. A vantagem deste processo é a redução dos possíveis efeitos causados pela seleção de indivíduos com maior *fitness*. A seleção por torneio, por sua vez, consiste em sortear aleatoriamente um grupo de indivíduos da população e, dentre estes, o que possuir melhor *fitness* é escolhido (KOZA, 1992). Recentemente o método do torneio tem se tornado um dos principais métodos de seleção devido a sua facilidade de ajustar a pressão de seleção, o que é feito apenas alterando o tamanho do grupo a ser sorteado (LI, 2000). Por essa razão, o método do torneio será utilizado em alguns dos experimentos descritos no capítulo 5.

É importante ressaltar que o indivíduo selecionado permanece na população original mesmo após a realização de qualquer operação. Dessa forma, é possível que um mesmo indivíduo seja escolhido mais de uma vez durante o processamento de uma geração (KOZA, 1992).

3.3.2 Reprodução

O operador de reprodução consiste simplesmente de copiar um indivíduo de uma geração para a próxima (KOZA, 1992; LI, 2000). Este mecanismo, que representa a base da seleção natural e sobrevivência do mais apto de Darwin, é executado basicamente em dois passos. No primeiro passo, escolhe-se um indivíduo da geração atual da população utilizando um método de seleção como os apresentados anteriormente. No segundo, copia-se este indivíduo para a nova geração, sem realizar qualquer modificação em sua estrutura (KOZA, 1992).

A grande vantagem deste método é o fato de o membro copiado entre gerações já ter sido avaliado em alguma interação anterior, removendo-se assim a necessidade de uma nova medição de *fitness* assim diminuindo a quantidade de processamento daquela execução em uma porcentagem aproximadamente igual à taxa de reprodução definida (KOZA, 1992).

3.3.3 Cruzamento

O cruzamento, ou *crossover*, consiste na troca de uma subárvore qualquer de um dado indivíduo por outra subárvore de outro indivíduo (SESHADRI, 2003) gerando, assim dois novos sujeitos de forma sexuada que serão inseridos na nova geração (KOZA, 1992). Esta operação tem início a partir da escolha de dois indivíduos da população original utilizando um método de seleção possivelmente igual ao utilizado para a operação de reprodução, os quais são chamados de pais. A seguir, escolhe-se de forma aleatória um ponto de cruzamento em cada um dos pais (LI, 2000). O resultado da escolha destes pontos são subárvores cujas raízes se encontram justamente no ponto selecionado. Com isso, para cada pai, é gerada uma nova árvore apenas removendo-se a referência para aquela subárvore indicada pelo ponto de cruzamento e inserindo, em seu lugar, a subárvore indicada pelo ponto de cruzamento do outro pai selecionado (KOZA, 1992). A Figura 3.3 ilustra de forma clara esta operação.

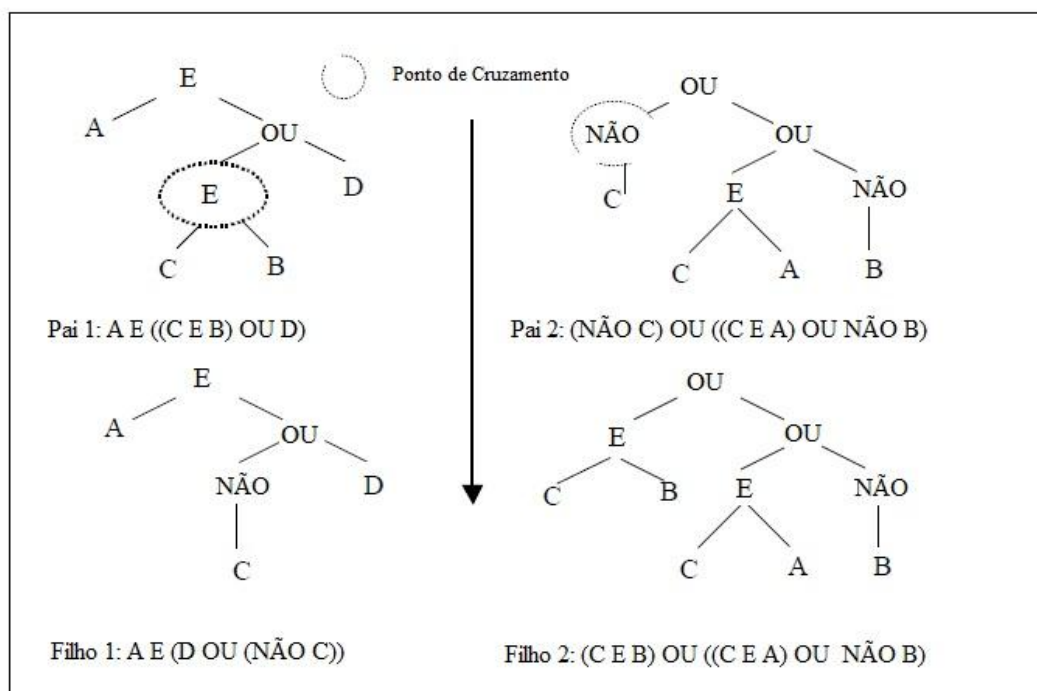


Figura 3.3: Operação de cruzamento mostrando pais e filhos gerados (LI, 2000).

É importante perceber que essa operação gera novas árvores que podem ter uma profundidade muito elevada. Para evitar que o tamanho exagerado das árvores geradas consuma muito tempo de processamento, pode-se definir um valor de altura máxima de forma que, se uma nova árvore ultrapassar esta altura, é escolhido um dos pais para ser reproduzido na nova geração. Nesse caso, se ambas ultrapassarem o limite de altura, os dois pais serão reproduzidos (KOZA, 1992).

Outro detalhe interessante sobre a operação de cruzamento em comparação com a operação de reprodução é que a segunda cria uma tendência em direção à convergência, enquanto a primeira gera uma pressão contrária a esta convergência fazendo com que um resultado único seja improvável (KOZA, 1992).

3.3.4 Mutação

Na área dos Algoritmos Genéticos, a operação de mutação é geralmente benéfica, pois introduz certa aleatoriedade em uma população, ajudando a escapar de mínimos locais em populações que estão convergindo rápido demais (KOZA, 1992).

Resumidamente, a operação de mutação é aplicada da seguinte forma (LI, 2000):

1. Seleciona-se um nodo aleatório da árvore.
2. Gera-se uma nova árvore.
3. Substitui-se a subárvore a partir do nodo selecionado pela nova árvore gerada.

Portanto, pode-se dizer que da mesma forma como é feito nas operações descritas anteriormente, um indivíduo da população é escolhido através de um método de seleção e, neste indivíduo, é escolhido um nodo aleatório da árvore que o representa. Em seguida, toda a subárvore a partir deste nodo é substituída por outra gerada

aleatoriamente. Um parâmetro específico pode ser utilizado aqui para controlar a altura máxima da subárvore a ser gerada que, em geral, é o mesmo utilizado para descrever o tamanho máximo dos indivíduos gerados na primeira geração (KOZA, 1992). A Figura 3.4 ilustra o antes e o depois da operação de mutação em uma árvore,

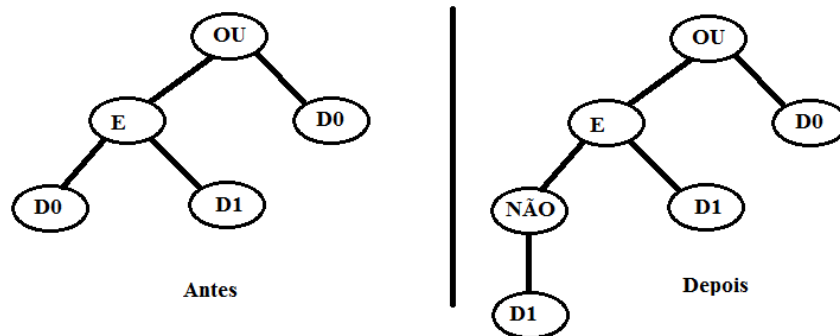


Figura 3.4: Operação de mutação com estado da árvore antes e depois (KOZA, 1992).

3.4 Critérios de Parada

Por fim, é importante saber determinar quando um resultado alcançado é bom o suficiente. Existem problemas onde se pode facilmente definir quando uma solução é a ideal e quando não é. Todavia, para muitos outros problemas, essa facilidade no reconhecimento da solução não existe (KOZA, 1992). Na programação genética, os problemas costumam se enquadrar no segundo caso. Dessa forma, define-se que a execução deve terminar quando atingir um número máximo de gerações ou quando o *fitness* considerado perfeito é alcançado (KOZA, 1992; LI, 2000). A partir disso usa-se o melhor indivíduo gerado até o momento como solução para o problema (KOZA, 1992).

4 MÉTODOS PARA PREVISÃO DE SÉRIES TEMPORAIS

A previsão dos valores futuros de algumas séries temporais se faz bastante necessária em várias áreas da indústria e economia nos tempos atuais. É muito comum encontrar casos onde seja necessário fazer previsões destas séries em intervalos pelo menos mensais (HYNDMAN; KHANDAKAR, 2008). Pensando nesta necessidade, nas últimas décadas foram criadas várias técnicas de previsão como, por exemplo, o ARIMA e as técnicas baseadas em Redes Neurais (SHABRI, 2001). Todavia, percebe-se que a falta de pessoas capacitadas para utilizar tais modelos é um grande problema desta área. Como solução para isso, criam-se ferramentas que utilizam estes modelos de forma a facilitar as previsões para o usuário final (HYNDMAN; KHANDAKAR, 2008).

Os métodos existentes para realizar as previsões de tais séries são classificados como quantitativos e baseiam-se na premissa de que os dados observados em um determinado tempo no passado definem o padrão de comportamento da série e este padrão tende a se repetir com o tempo (BECKER, 2010). Portanto, neste capítulo serão apresentadas, individualmente, os quatro métodos comparados neste trabalho de forma a prover um conhecimento mais amplo sobre os funcionamentos de cada técnica utilizada.

4.1 Naïve

O primeiro método de previsão de séries temporais também é o mais simples de todos e é, geralmente, utilizado como base para comparação com outras técnicas. Este método consiste de apenas repetir o uso de determinado valor para realizar a previsão dos valores futuros (CARBONNEAU; LAFRAMBOISE; VAHIDOV, 2008).

Existem variações mais utilizadas deste método, de acordo com (LEICHT; TANNER, 1991):

- **Naïve sem mudanças:** esta variação consiste de simplesmente repetir o último valor mensurado da série para todos os valores futuros desejados pelo usuário.
- **Naïve com variação constante:** esta variante, por sua vez, é um pouco mais elaborada e assume que a variação para os próximos 'x' intervalos futuros é a mesma variação medida para o último intervalo.

As figuras 4.1 e 4.2 mostram um exemplo de previsão utilizando cada um dos métodos citados, onde o eixo horizontal é uma medida de tempo, o vertical é o valor correspondente àquele instante e as cores indicam os intervalos de confiança para as previsões. Baseando-se nestas figuras, verifica-se que o gráfico resultante de uma previsão feita pelo método **Naïve sem mudanças** é uma reta horizontal e pelo método

Naïve com variação constante é uma reta cuja inclinação é igual à variação entre o penúltimo e o último resultado medido.

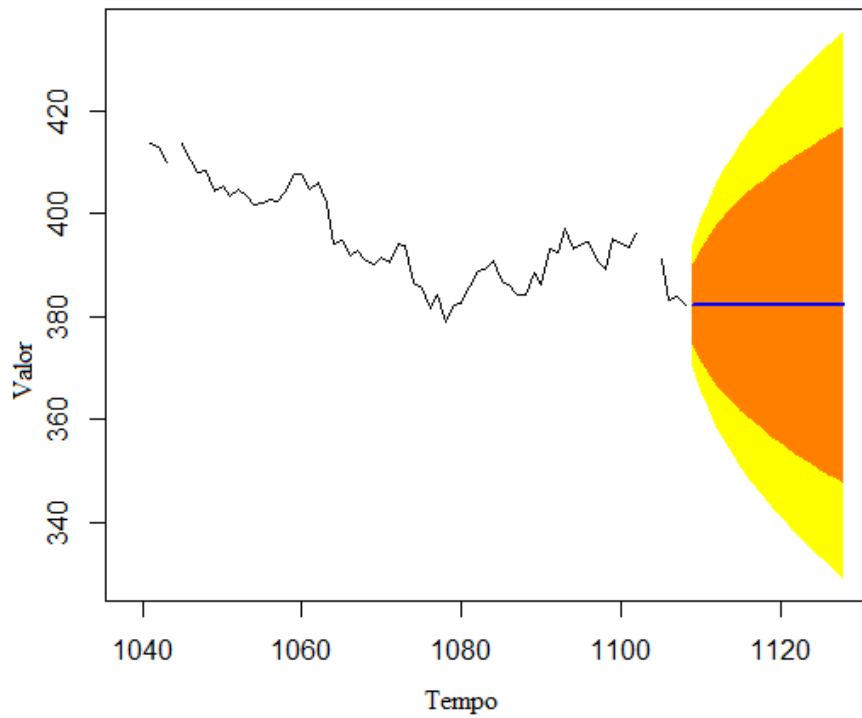


Figura 4.1: Gráfico resultante da previsão de uma série utilizando o método Naïve sem mudanças.

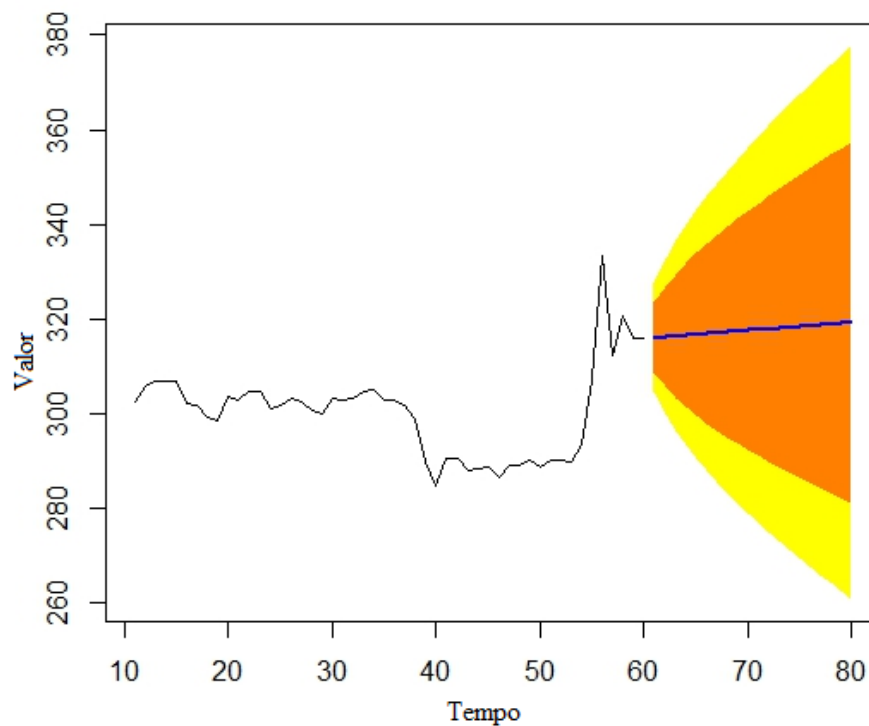


Figura 4.2: Gráfico resultante da previsão de uma série utilizando o método Naïve com variação constante.

4.2 ARIMA

O método ARIMA (*Auto Regressive Integrated Moving Averages*), também conhecido como método Box-Jenkins, utiliza um algoritmo complexo baseado em médias móveis e termos auto regressivos para definir o modelo matemático mais apropriado para representar o comportamento da série temporal estudada baseando-se na auto-correlação entre os valores da série (FERNANDES, 2010). Sendo um dos métodos mais utilizados na prática para a previsão de séries temporais, o ARIMA relaciona valores atuais de uma série não só com os valores históricos desta, mas também com os abalos aleatórios ocorridos no passado e os que ocorrem no momento (KLÚČIK; JURIOVÁ; KLÚČIK, 2009).

$$\Phi_p(B)\Phi_p(B^L)(1-B^L)^D(1-B)^d y_t = a + \Theta_q(B)\Theta_q(B^L)\varepsilon_t \quad (1)$$

Mais especificamente, o modelo de um processo ARIMA (p, d, q) pode ser representado pela equação na equação (1), onde temos (SHABRI, 2001; HYNDMAN; KHANDAKAR, 2008):

- **p**: grau do polinômio do fator auto regressivo;
- **d**: grau da tendência de variação da série;
- **q**: grau do polinômio da Média Móvel;
- $\Phi(B)$: é o operador auto regressivo e um polinômio de **grau p**;
- $\Theta(B)$: é o operador da média móvel e um polinômio de **grau q**;
- **B**: é o operador de movimentação na série
- ε_t : é um erro aleatório com distribuição normal $N(0, \sigma^2)$;
- **a**: é uma constante;
- y_t : são os dados da série temporal.

A previsão dos valores futuros da série utilizando este método segue quatro etapas principais (RATNIEKS, 2010). No primeiro passo, deve-se identificar um modelo ARIMA apropriado para a série. Em seguida, é necessário identificar os parâmetros apropriados para o modelo escolhido. Feito isso, são executados alguns diagnósticos sobre o modelo escolhido e, se necessário, este é alterado. Por fim, na quarta etapa, o melhor modelo escolhido nos passos anteriores é utilizado para a previsão em si (SHABRI, 2001). A Figura 4.3 exemplifica a previsão de uma série temporal utilizando o ARIMA.

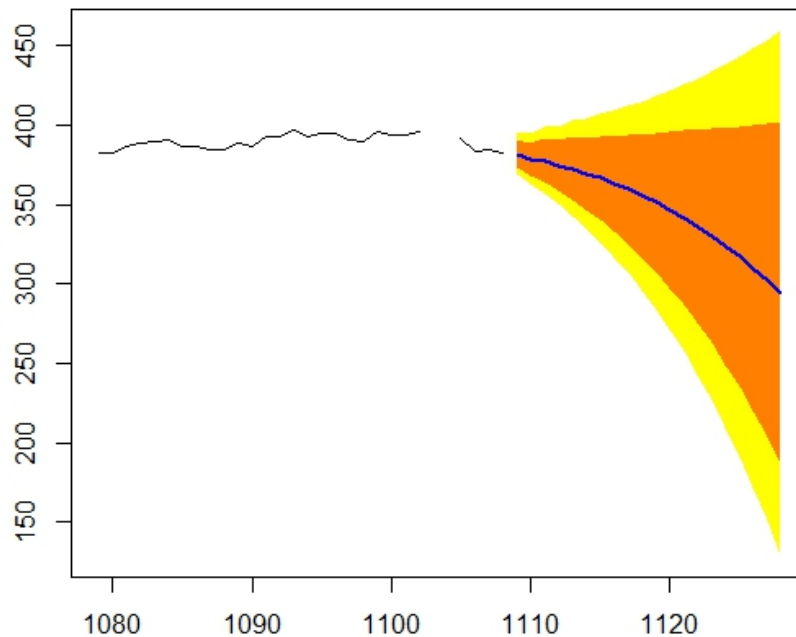


Figura 4.3: Exemplo de previsão utilizando ARIMA (3, 4, 5).

4.3 Redes Neurais

As Redes Neurais são estruturas artificiais inspiradas no funcionamento do cérebro humano (SHABRI, 2001). Atualmente a qualidade dos resultados utilizando Redes Neurais desafia até os melhores dos algoritmos clássicos de previsão de séries temporais. Nesse contexto, tornou-se muito popular a utilização do Perceptron Multicamadas para a realização destas previsões (GHEYAS; SMITH, 2009).

Em geral, uma Rede Neural possui uma estrutura como a apresentada na Figura 4.4, contendo três camadas interconectadas através de conexões ponderadas (SHABRI, 2001). A camada de entrada apenas recebe as N medições da série analisada realizadas até um ponto t e repassa para os neurônios da camada oculta que, por sua vez, realizam somas ponderadas de determinados valores recebidos e repassam o resultado para a camada de saída. Finalmente, a camada de saída realiza outra soma ponderada dos resultados obtidos da camada oculta e gera a previsão para o próximo valor da série (FRANK; DAVEY; HUNT, 2001). A Figura 4.5 ilustra a passagem de informação entre as camadas de forma a gerar a previsão do próximo valor da série.

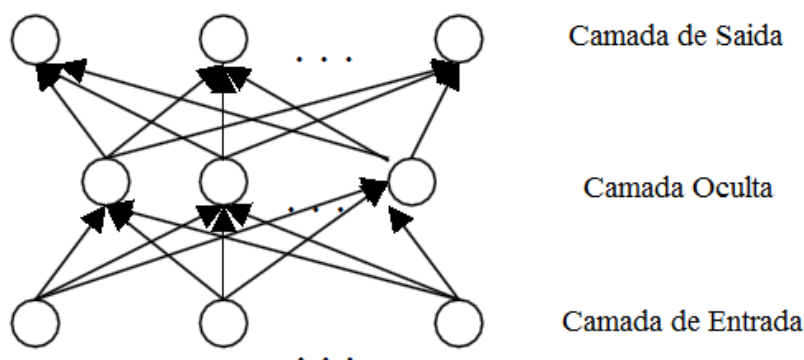


Figura 4.4: Rede Neural Típica (SHABRI 2001).

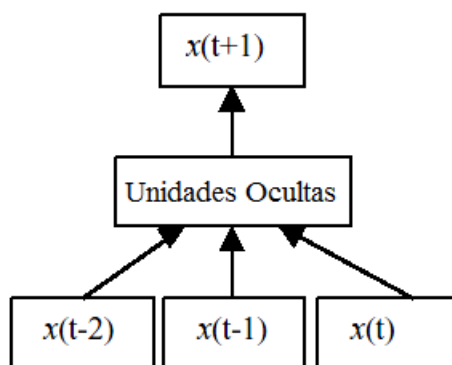


Figura 4.5: Fluxo para a geração da previsão do próximo valor de uma série (FRANK; DAVEY; HUNT, 2001).

A utilização das N medições da série temporal até um ponto é chamada de **Técnica da Janela Deslizante** uma vez que durante a fase de treinamento da Rede Neural deve-se passar por todos os pontos do conjunto de treinamento utilizando os N pontos anteriores ao ponto de teste. Portanto, a variação no tamanho da janela tem um grande impacto na qualidade das previsões feitas pela rede (FRANK; DAVEY; HUNT, 2001). Ainda durante a fase de treinamento, os pesos das conexões entre os neurônios vão se alterando de acordo com uma regra de aprendizado de forma a diminuir o erro das previsões realizadas pela rede (SHABRI, 2001).

Por fim, estudos indicam que os resultados das previsões utilizando Redes Neurais são muito promissores, ainda mais quando comparados com outros métodos como o ARIMA, onde as redes podem apresentar resultados de igual ou até melhor qualidade (SHABRI, 2001). A título de exemplo, a Figura 4.6 mostra o resultado da previsão de valores de uma série temporal utilizando Redes Neurais.

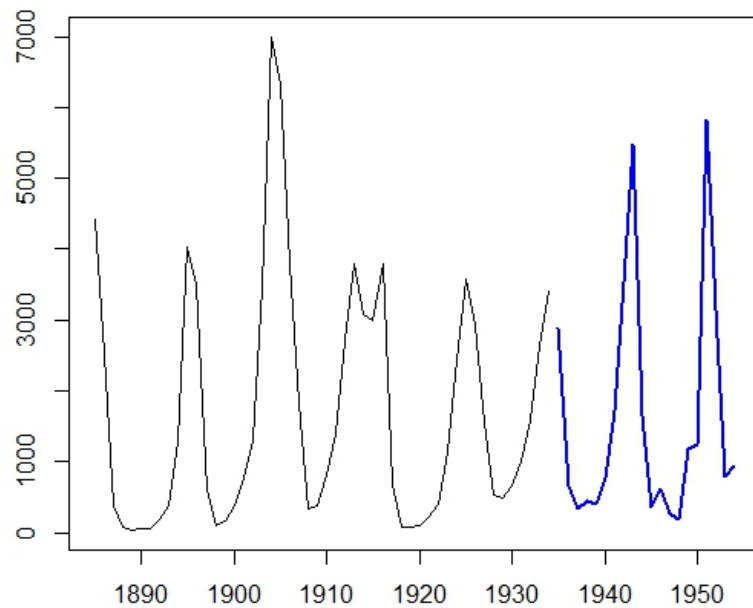


Figura 4.6: Exemplo de previsão de uma série temporal utilizando Redes Neurais.

4.4 Regressão Simbólica

Como foi apresentado no capítulo 3, a Programação Genética foi criada baseada nos conceitos do processo evolutivo existentes nos Algoritmos Genéticos. Com essa técnica, é possível realizar o processo denominado Regressão Simbólica (KLÚČIK; JURIOVÁ; KLÚČIK, 2009).

A Regressão Simbólica, também conhecida como Procura de Funções, é um processo de indução em expressões matemáticas baseando-se em dados de um sistema, como uma série, por exemplo, para realizar alterações nas estruturas destas expressões. Nesse processo, procura-se encontrar uma relação funcional f tal que $f: R_n \rightarrow R$, ou seja, para um dado conjunto de entradas, seja possível produzir valores de saída que se aproximem ao máximo dos valores considerados no sistema (GRINGS, 2006).

Este método, por se tratar de Programação Genética, requer que parâmetros específicos tais como o conjunto de funções e terminais, tamanho da população, número máximo de gerações e outros citados no capítulo 3 tenham de ser informados. Além disso, em geral, utiliza-se como função de *fitness* o erro calculado entre o valor da série e o valor retornado pela função (KLÚČIK; JURIOVÁ; KLÚČIK, 2009).

A utilização deste método apresenta algumas vantagens em relação a outros métodos. Por exemplo, quando se utiliza Redes Neurais ou o próprio ARIMA, é muito difícil ou até impossível verificar qual a função que descreve o comportamento da aproximação. Em Programação Genética, por outro lado, para obter esta função basta verificar a árvore que descreve o indivíduo escolhido ao final da execução (KLÚČIK; JURIOVÁ; KLÚČIK, 2009; GRINGS, 2006). Outra grande vantagem da Regressão Simbólica está na facilidade de paralelização da Programação Genética. Como a tarefa que mais requer poder computacional é a avaliação e esta pode ser executada apenas com as informações do indivíduo e da série considerada, fica fácil realizar a execução de uma forma paralela (GRINGS, 2006).

Estudos indicam que a utilização da Regressão Simbólica é capaz de produzir resultados de qualidade equivalente aos encontrados quando utilizando o método ARIMA, mas com as vantagens já explicadas anteriormente (KLÚČIK; JURIOVÁ; KLÚČIK, 2009). Por essa razão, esse trabalho faz experimentos com essa técnica a fim de comparar seus resultados com outros obtidos através da utilização de outros métodos clássicos.

5 PROCEDIMENTO EXPERIMENTAL

Nos capítulos anteriores foi feita uma breve apresentação de alguns conceitos muito importantes para o entendimento do experimento como um todo. Neste capítulo serão apresentados e discutidos os resultados do processamento de dados de séries temporais produzidas nos mercados de ações e Forex, apresentados no capítulo 2, através dos métodos de previsão de séries discutidos no capítulo 4. Primeiramente, será apresentada a forma como os dados foram coletados, seguido das ferramentas e parâmetros utilizados para as simulações e, finalmente, os resultados obtidos por cada método ao serem aplicados em uma mesma série temporal. Na última seção deste capítulo serão discutidas as diferenças encontradas nos experimentos.

5.1 Coleta e Organização dos Dados

Para a execução dos experimentos, foram escolhidas duas séries geradas no Mercado de Ações e duas séries geradas no Mercado Forex de forma que os testes pudessem ser feitos em dados gerados em datas e locais diferentes. Do Mercado de Ações foram obtidas as informações, medidas diariamente, referentes às ações da empresa Petrobrás (PETR4) e da Companhia Vale do Rio Doce (VALE5), disponibilizadas em (UOL ECONOMIA, 2012). Do Mercado Forex, por sua vez, foram obtidas as informações referentes à variação diária do Dólar Americano em relação ao Euro (EURUSD) e em relação à Libra Britânica (GBPUSD), disponibilizadas em (FX HISTORICAL DATA, 2012). Como explicado em seções anteriores, estas séries específicas foram escolhidas devido às suas popularidades em seus respectivos mercados.

Obtidos os dados, foi necessário escolher um subconjunto de cada série para realizar o processamento. Dessa forma, utilizaram-se os dados entre 26 de agosto de 2011 e 20 de setembro de 2012 para as séries do Mercado Forex. As informações das ações da PETR4 utilizadas estão datadas entre 16 de novembro de 2006 e 7 de maio de 2008, enquanto que as da VALE5 estão entre 2 de janeiro de 2009 e 15 de junho de 2010.

5.2 O Projeto R

O R é uma ferramenta feita para ser utilizada em tarefas de estatística, gráficos e programação estatística (TEETOR, 2011). Na última década esta ferramenta vem se desenvolvendo muito e hoje ela é utilizada em várias grandes empresas como a Microsoft, o Google, a Pfizer, entre outras (ADLER, 2012). Além de ser uma ferramenta gratuita, ela conta com a colaboração de muitos desenvolvedores ao redor do mundo e uma vasta gama de pacotes que podem ser instalados também gratuitamente (HORTON; KLEINMAN, 2011).

Devido ao fato desta ferramenta possuir formas bastante intuitivas de gerar gráficos e algumas previsões, ela foi utilizada para aplicar os métodos de predição de séries temporais descritos no capítulo 4. Considerando que a apresentação detalhada da ferramenta não é o foco deste trabalho sugere-se a leitura da documentação da mesma, disponível em (RPROJECT, 2012), para maiores detalhes sobre seu funcionamento.

O Projeto R (RPROJECT, 2012) juntamente com a ferramenta R Studio (RSTUDIO, 2012) foram utilizados neste trabalho para realizar as simulações dos métodos **Naïve**, **ARIMA** e **Redes Neurais** e gerar os gráficos referentes a cada uma das execuções. Como o método **ARIMA** é o único que depende de parâmetros para ser executado, após uma serie de testes com diferentes combinações de valores, foi definido que os valores utilizados para **p**, **d** e **q** seriam, respectivamente, 3, 4 e 9 para todas as séries consideradas. Na sessão 5.4 serão apresentados todos os resultados obtidos ao aplicar cada um dos métodos de predição expostos no capítulo 4 aos dados selecionados e mencionados na sessão 5.1.

5.3 O ECJ e a Regressão Simbólica

O ECJ, disponível em (LUKE et al., 2012), é uma ferramenta para a criação de programas que utilizam a Computação Evolutiva como base. Sendo assim, utilizou-se o ECJ com a finalidade de realizar a predição das séries temporais através da Regressão Simbólica. O ECJ é uma ferramenta de código aberto que, ao longo de mais de dez anos, acabou se tornando muito estável. Foi desenvolvida em Java e hoje é utilizada para experimentos que requerem o processamento de uma carga pesada de dados (LUKE, 2010).

Além disso, o ECJ provê uma série de características, como otimizações com multi objetivos e a própria implementação de Programação Genética, que visam facilitar o trabalho do programador (LUKE, 2012). Adicionalmente, existem alguns exemplos de aplicações que são providas pelo próprio ECJ e apresentam os fundamentos da ferramenta ao programador, uma vez que o seu aprendizado pode ser bastante difícil (LUKE, 2010).

Da mesma forma como foi explicado na sessão 5.2, o foco deste trabalho não é a apresentação da ferramenta. Sendo assim, caso haja interesse em detalhes específicos sobre o ECJ, sugere-se a leitura do manual da ferramenta em (LUKE, 2010).

5.3.1 Histórico do Experimento utilizando ECJ

Para a realização do experimento utilizando Programação Genética e a Regressão Simbólica, explicada na sessão 4.4, é necessário definir uma série de detalhes, como foi visto no capítulo 3. Para isso, foram feitos uma série de experimentos variando diversas combinações de parâmetros de forma a escolher os que melhor satisfiziam o problema em um tempo aceitável.

Iniciaram-se os testes utilizando as séries selecionadas contendo aproximadamente três mil pontos cada uma, e com a seguinte configuração para a Programação Genética:

- **Conjunto de Funções:** {Soma, Subtração, Multiplicação, Divisão, Seno, Cosseno, Potenciação, Exponenciação, Logaritmo};

- **Conjunto de Terminais:** {T (variável de tempo), Constante};
- **Número Máximo de Gerações:** 10000;
- **Função de Fitness:** RMSE;
- **Tamanho da População:** 1000;
- **Altura Máxima da Árvore:** 22;
- **Número de Seleccionados por Elitismo:** 3.

Com isso, verificou-se que a execução do algoritmo requeria muito tempo de processamento, chegando a demorar mais de um dia e meio para aproximar uma só série, mesmo com um erro bastante elevado. A partir deste resultado, percebeu-se que alguns detalhes poderiam ser modificados para acelerar a execução do algoritmo. A principal medida foi reduzir a quantidade de pontos da Serie Temporal utilizada, uma vez que o método precisava, ao realizar as avaliações, passar por todos os pontos desta série, para cada indivíduo e a cada geração.

Em seguida, tentou-se remover todas aquelas funções que dependiam de uma única variável, mas aumentando a altura máxima da árvore. Com essas medidas, foi possível obter uma qualidade consideravelmente melhor na aproximação da série original e uma redução no número de gerações requeridas para obter este resultado, diminuindo, portanto, o tempo total de execução do algoritmo para, em média, 5 horas. Por fim, os parâmetros utilizados para os testes com as quatro Séries Temporais coletados foram os seguintes:

- **Conjunto de Funções:** {Soma, Subtração, Multiplicação, Divisão};
- **Conjunto de Terminais:** {T (variável de tempo), Constante};
- **Número Máximo de Gerações:** 2500;
- **Função de Fitness:** RMSE;
- **Tamanho da População:** 700;
- **Altura Máxima da Árvore:** 27;
- **Total de Participantes do Torneio (Seleção):** 7;
- **Número de Seleccionados por Elitismo:** 3.

Ainda nesses experimentos foi aproveitada a alta capacidade de paralelização da Programação Genética, utilizando múltiplas *threads* para realizar as avaliações e

seleções na população. Ao fim da execução, o ECJ disponibiliza um arquivo que contém a melhor árvore encontrada para aproximar a Série Temporal, a qual pode ser utilizada para calcular os valores para diferentes valores de tempo. De posse destas árvores, foi possível calcular a previsão para os 30 próximos valores das séries testadas, sendo os gráficos destas previsões gerados com o auxílio da ferramenta R (RPROJECT, 2012).

É importante ressaltar que para cada série utilizada, a Regressão Simbólica foi executada cinco vezes, resultando em cinco diferentes funções de aproximação para a série original. Como resultado final, foi escolhida a função cujo valor de *fitness* era o melhor entre todos, fazendo com que esta função seja a que melhor aproxime a série testada.

5.4 Resultados dos Experimentos Realizados

Nesta seção, são apresentados os gráficos dos resultados obtidos através de experimentos com cada um dos métodos apresentados ao longo deste trabalho, para quatro Séries Temporais diferentes. Em todos os gráficos apresentados, a série original está colocada em preto e as previsões são colocadas em azul ou vermelho.

Para todos os métodos testados, foram previstos 30 pontos futuros e estes respectivos valores são apresentados sempre à direita da reta vertical em verde. Para o caso dos gráficos do método de Regressão Simbólica, os valores em vermelho à esquerda desta reta vertical representam o resultado da função de aproximação da série testada.

5.4.1 Resultados para EURUSD

As Figuras 5.1 a 5.4 representam os resultados obtidos quando os métodos apresentados foram aplicados à série EURUSD. Nos gráficos, os pontos entre 0 e 250 representam valores conhecidos em tempo de execução ou treinamento dos métodos e os valores entre 251 e 280 são as tentativas de previsão realizadas.

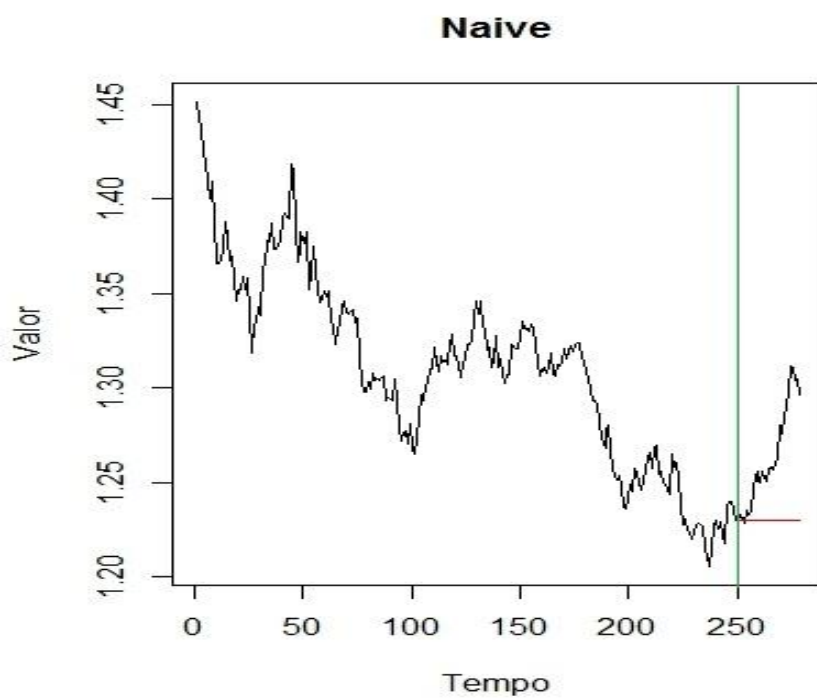


Figura 5.1: Previsão dos 30 próximos valores da série EURUSD usando Naive.

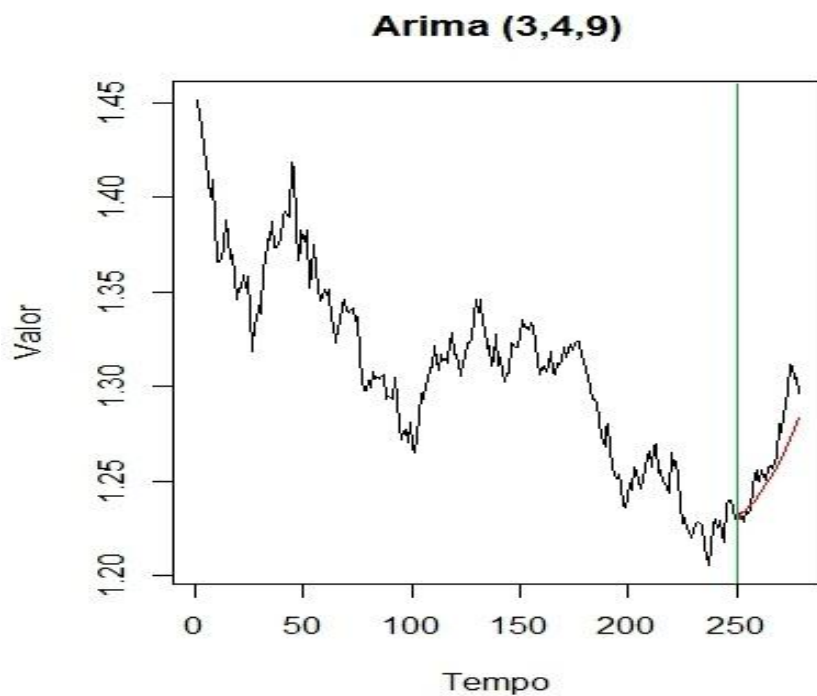


Figura 5.2: Previsão dos 30 próximos valores da série EURUSD usando ARIMA.

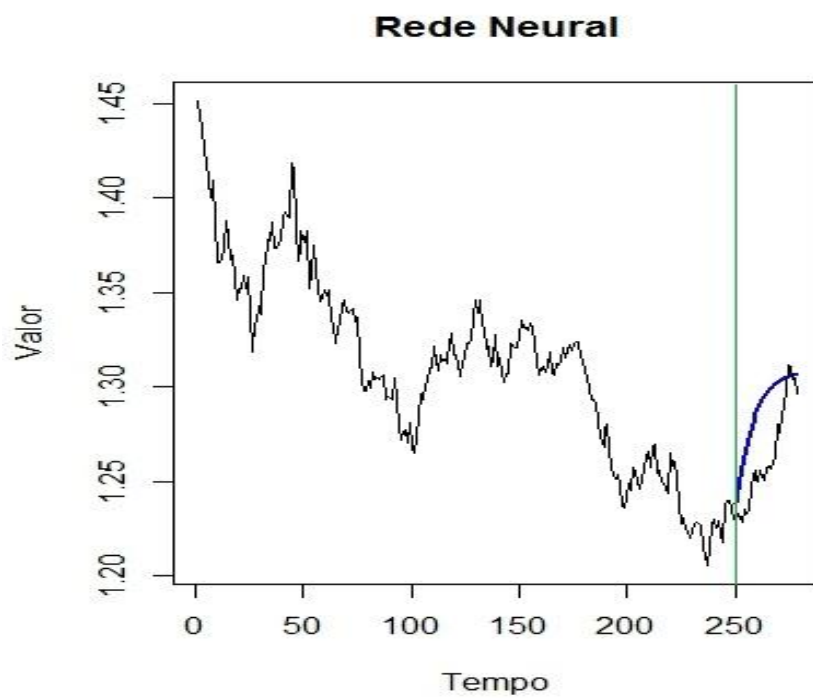


Figura 5.3: Previsão dos 30 próximos valores da série EURUSD usando Redes Neurais.

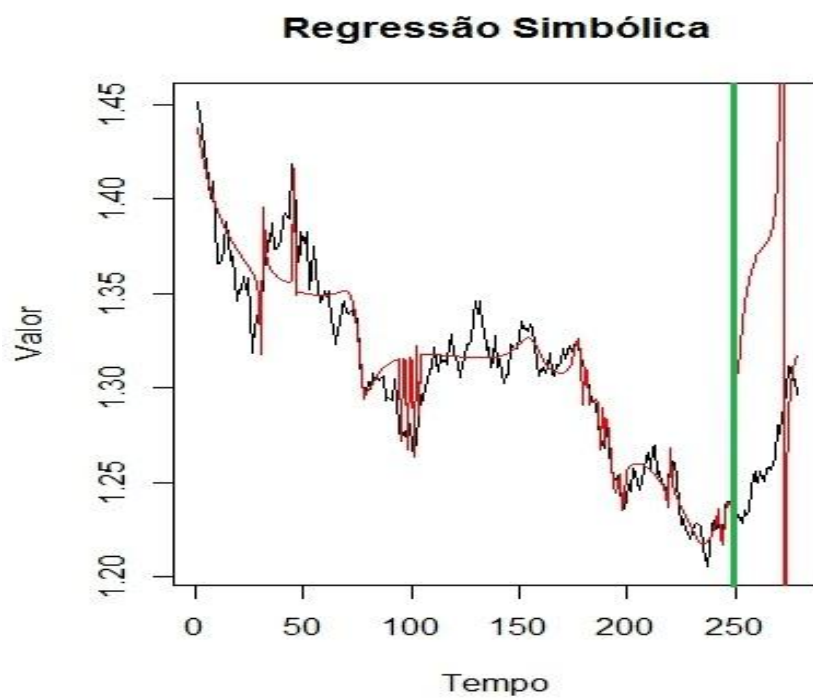


Figura 5.4: Previsão dos 30 próximos valores da série EURUSD usando Regressão Simbólica.

Para os 30 valores previstos em cada método, a Tabela 5.1 apresenta as diferenças quadráticas entre o valor real e o valor previsto, além do RMSE das previsões realizadas por cada método.

Tabela 5.1: Valores reais, valores previstos e erros nas previsões para a série EURUSD

EURUSD									
#	Real	Naïve		ARIMA		Rede Neural		Regressão Simbólica	
		Previsto	Erro Quadrático	Previsto	Erro Quadrático	Previsto	Erro Quadrático	Previsto	Erro Quadrático
1	1,2291	1,2302	0,000001	1,2276	0,000002	1,2381	0,000080	1,3054	0,005823
2	1,2330	1,2302	0,000008	1,2306	0,000006	1,2453	0,000151	1,3093	0,005829
3	1,2316	1,2302	0,000002	1,2337	0,000004	1,2519	0,000411	1,3207	0,007934
4	1,2284	1,2302	0,000003	1,2341	0,000032	1,2579	0,000868	1,3346	0,011278
5	1,2355	1,2302	0,000028	1,2345	0,000001	1,2633	0,000773	1,3433	0,011632
6	1,2332	1,2302	0,000009	1,2370	0,000014	1,2682	0,001227	1,3510	0,013882
7	1,2344	1,2302	0,000018	1,2367	0,000005	1,2727	0,001464	1,3570	0,015026
8	1,2468	1,2302	0,000276	1,2382	0,000074	1,2767	0,000892	1,3615	0,013156
9	1,2528	1,2302	0,000511	1,2394	0,000180	1,2803	0,000754	1,3649	0,012565
10	1,2562	1,2302	0,000676	1,2410	0,000232	1,2835	0,000744	1,3674	0,012374
11	1,2510	1,2302	0,000433	1,2416	0,000089	1,2864	0,001250	1,3694	0,014011
12	1,2495	1,2302	0,000372	1,2436	0,000034	1,2889	0,001555	1,3709	0,014734
13	1,2563	1,2302	0,000681	1,2448	0,000131	1,2912	0,001220	1,3722	0,013422
14	1,2528	1,2302	0,000511	1,2465	0,000040	1,2933	0,001638	1,3733	0,014529
15	1,2504	1,2302	0,000408	1,2479	0,000006	1,2951	0,001997	1,3746	0,015428
16	1,2575	1,2302	0,000745	1,2500	0,000056	1,2967	0,001537	1,3762	0,014083
17	1,2585	1,2302	0,000801	1,2515	0,000049	1,2981	0,001571	1,3783	0,014353
18	1,2565	1,2302	0,000692	1,2536	0,000008	1,2994	0,001842	1,3814	0,015609
19	1,2600	1,2302	0,000888	1,2555	0,000020	1,3005	0,001644	1,3864	0,015965
20	1,2630	1,2302	0,001076	1,2577	0,000028	1,3015	0,001485	1,3947	0,017339
21	1,2802	1,2302	0,002500	1,2597	0,000420	1,3024	0,000494	1,4106	0,017006
22	1,2757	1,2302	0,002070	1,2621	0,000184	1,3032	0,000757	1,4495	0,030212
23	1,2850	1,2302	0,003003	1,2644	0,000425	1,3039	0,000357	1,6513	0,134146
24	1,2896	1,2302	0,003528	1,2669	0,000515	1,3045	0,000222	1,0518	0,056543
25	1,2984	1,2302	0,004651	1,2694	0,000843	1,3050	0,000044	1,2464	0,002703
26	1,3117	1,2302	0,006642	1,2721	0,001567	1,3055	0,000038	1,2848	0,000725
27	1,3112	1,2302	0,006561	1,2748	0,001327	1,3059	0,000028	1,3006	0,000113
28	1,3042	1,2302	0,005476	1,2777	0,000704	1,3063	0,000004	1,3092	0,000025
29	1,3044	1,2302	0,005506	1,2806	0,000568	1,3066	0,000005	1,3139	0,000091
30	1,2964	1,2302	0,004382	1,2837	0,000162	1,3069	0,000111	1,3170	0,000426
		RMSE=	0,041816		0,016051		0,028962		0,129223

5.4.2 Resultados para GBPUSD

Seguindo o mesmo formato apresentado na sessão anterior, as Figuras 5.5 a 5.8 representam os resultados obtidos quando os métodos apresentados foram aplicados à série GBPUSD. Nestes gráficos, os pontos de Tempo de 0 a 250 representam valores conhecidos quando executados os métodos de predição e os pontos entre 251 e 280 são as previsões feitas.

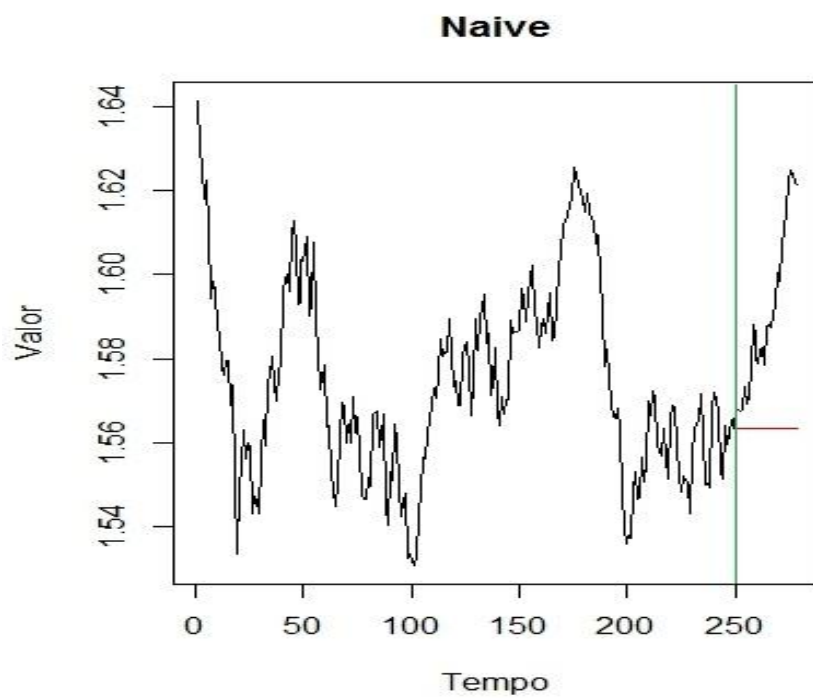


Figura 5.5: Previsão dos 30 próximos valores da série GBPUSD usando Naive.

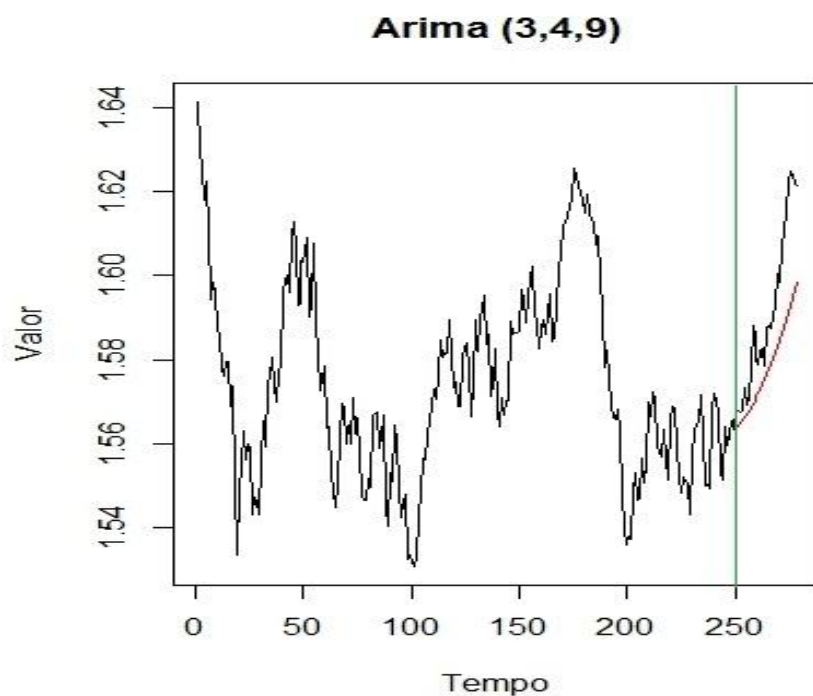


Figura 5.6: Previsão dos 30 próximos valores da série GPBUSD usando ARIMA.

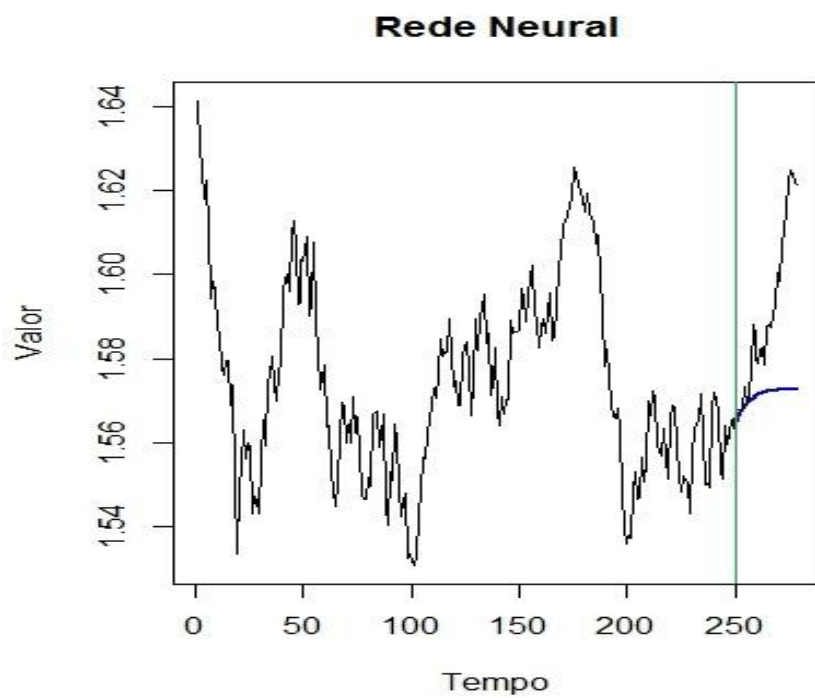


Figura 5.7: Previsão dos 30 próximos valores da série GBPUSD usando Redes Neurais.

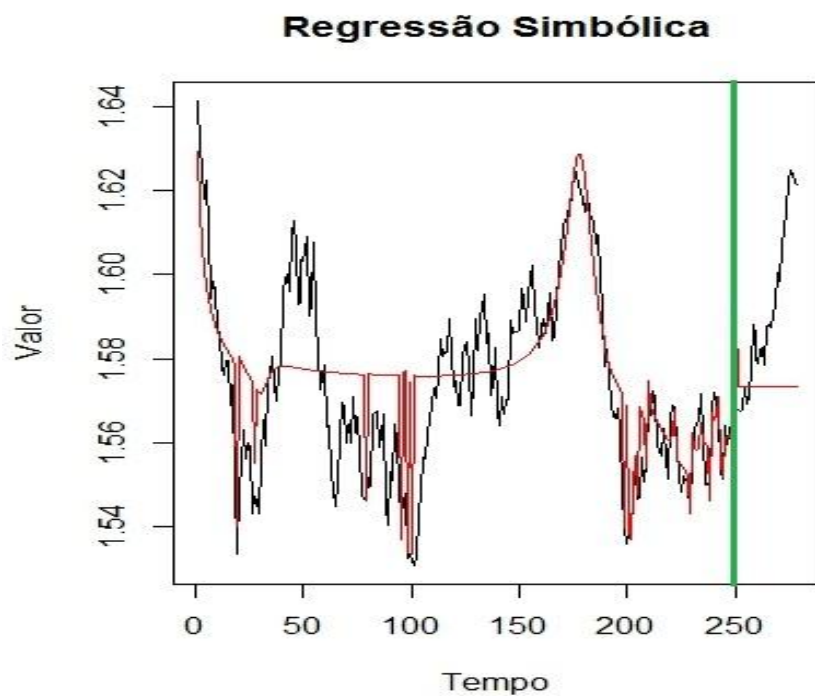


Figura 5.8: Previsão dos 30 próximos valores da série GBPUSD usando Regressão Simbólica.

A Tabela 5.2 apresenta os valores de cada ponto previsto, além da diferença quadrática na previsão destes e do RMSE para os valores previstos da série GBPUSD, considerando cada método separadamente.

Tabela 5.2: Valores reais, valores previstos e erros nas previsões para a série GBPUSD

GBPUSD									
#	Real	Naïve		ARIMA		Rede Neural		Regressão Simbólica	
		Previsto	Erro Quadrático	Previsto	Erro Quadrático	Previsto	Erro Quadrático	Previsto	Erro Quadrático
1	1,5681	1,5633	0,000023	1,5636	0,000020	1,5647	0,000011	1,5915	0,000550
2	1,5679	1,5633	0,000021	1,5642	0,000014	1,5660	0,000004	1,5732	0,000028
3	1,5674	1,5633	0,000017	1,5651	0,000005	1,5670	0,000000	1,5733	0,000035
4	1,5678	1,5633	0,000020	1,5653	0,000006	1,5679	0,000000	1,5733	0,000030
5	1,5734	1,5633	0,000102	1,5662	0,000052	1,5687	0,000023	1,5733	0,000000
6	1,5691	1,5633	0,000034	1,5670	0,000004	1,5693	0,000000	1,5733	0,000018
7	1,5708	1,5633	0,000056	1,5676	0,000010	1,5698	0,000001	1,5733	0,000006
8	1,5781	1,5633	0,000219	1,5685	0,000093	1,5703	0,000061	1,5733	0,000023
9	1,5880	1,5633	0,000610	1,5693	0,000349	1,5707	0,000300	1,5733	0,000216
10	1,5856	1,5633	0,000497	1,5702	0,000238	1,5710	0,000213	1,5733	0,000152
11	1,5801	1,5633	0,000282	1,5711	0,000081	1,5713	0,000078	1,5733	0,000046
12	1,5788	1,5633	0,000240	1,5721	0,000045	1,5715	0,000053	1,5733	0,000030
13	1,5818	1,5633	0,000342	1,5731	0,000076	1,5717	0,000102	1,5733	0,000072
14	1,5827	1,5633	0,000376	1,5742	0,000073	1,5719	0,000117	1,5733	0,000089
15	1,5783	1,5633	0,000225	1,5753	0,000009	1,5720	0,000039	1,5733	0,000025
16	1,5873	1,5633	0,000576	1,5764	0,000118	1,5721	0,000230	1,5733	0,000196
17	1,5882	1,5633	0,000620	1,5777	0,000111	1,5722	0,000255	1,5733	0,000222
18	1,5873	1,5633	0,000576	1,5789	0,000070	1,5723	0,000224	1,5733	0,000196
19	1,5900	1,5633	0,000713	1,5803	0,000095	1,5724	0,000310	1,5733	0,000280
20	1,5933	1,5633	0,000900	1,5816	0,000136	1,5725	0,000434	1,5733	0,000401
21	1,6005	1,5633	0,001384	1,5830	0,000305	1,5725	0,000784	1,5733	0,000741
22	1,5985	1,5633	0,001239	1,5845	0,000195	1,5726	0,000673	1,5733	0,000637
23	1,6066	1,5633	0,001875	1,5861	0,000422	1,5726	0,001157	1,5733	0,001111
24	1,6106	1,5633	0,002237	1,5877	0,000526	1,5726	0,001443	1,5733	0,001395
25	1,6148	1,5633	0,002652	1,5893	0,000650	1,5726	0,001777	1,5732	0,001727
26	1,6224	1,5633	0,003493	1,5910	0,000985	1,5727	0,002473	1,5732	0,002417
27	1,6246	1,5633	0,003758	1,5928	0,001012	1,5727	0,002695	1,5732	0,002640
28	1,6240	1,5633	0,003684	1,5946	0,000864	1,5727	0,002631	1,5732	0,002580
29	1,6219	1,5633	0,003434	1,5965	0,000645	1,5727	0,002419	1,5732	0,002372
30	1,6213	1,5633	0,003364	1,5984	0,000522	1,5727	0,002359	1,5732	0,002316
		RMSE=	0,033452		0,016054		0,026374		0,026174

5.4.3 Resultados para PETR4

A primeira das séries utilizadas do Mercado de Ações foi a PETR4. As Figuras 5.9 a 5.12 mostram os resultados das simulações com esta Série Temporal. É importante ressaltar que para estes testes foi utilizada uma série um pouco maior a fim de poder testar a validade dos métodos com diferentes tamanhos de séries. Dessa forma, os valores de Tempo entre 0 e 330 eram conhecidos dos algoritmos em tempo de treinamento e os valores de 331 a 360 foram previsões realizadas pelos métodos.

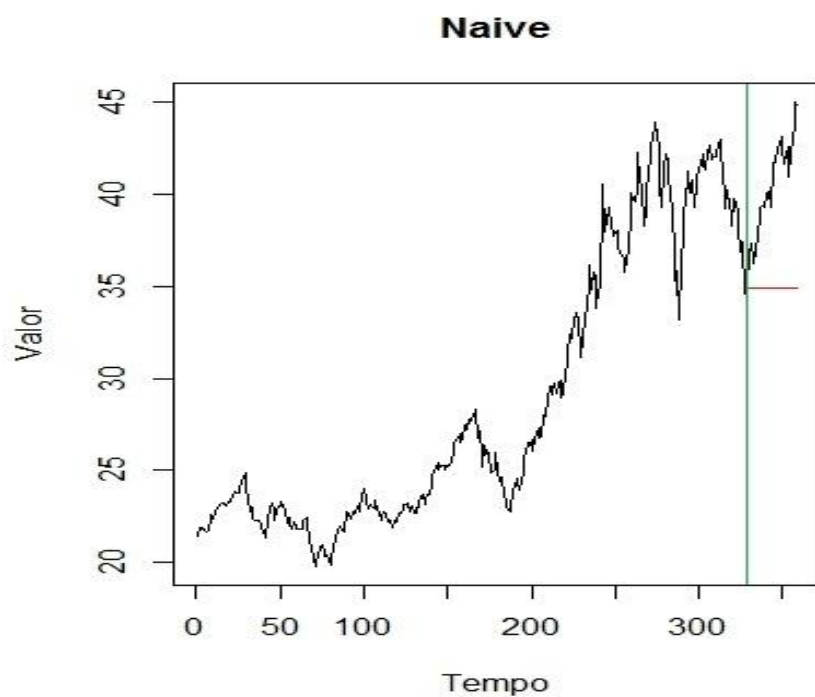


Figura 5.9: Previsão dos 30 próximos valores da série PETR4 usando Naïve.

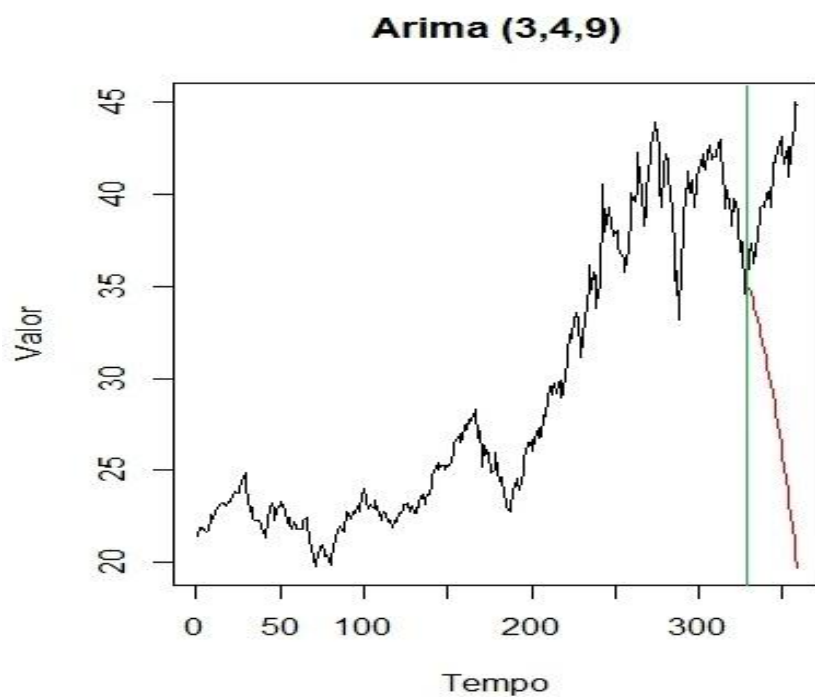


Figura 5.10: Previsão dos 30 próximos valores da série PETR4 usando ARIMA.

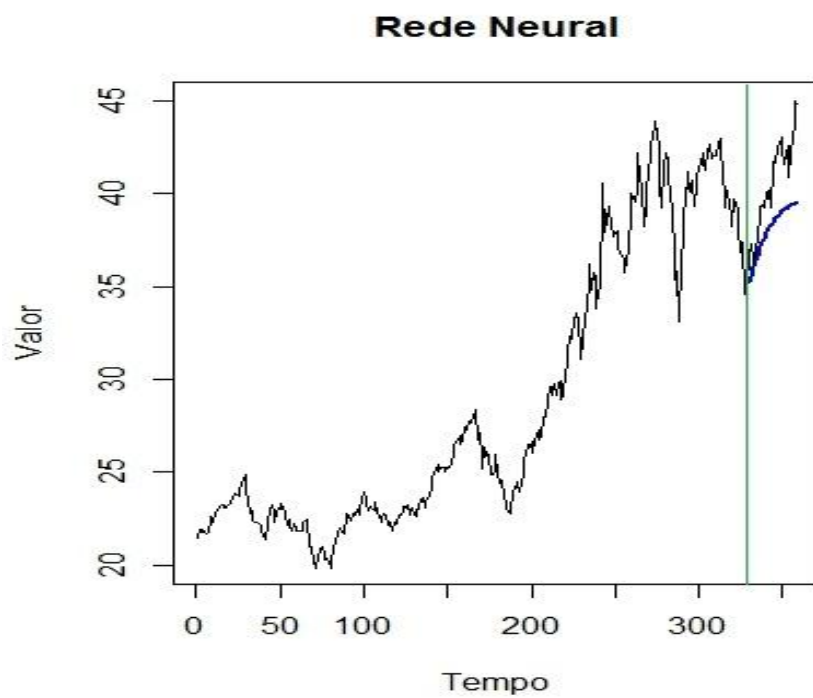


Figura 5.11: Previsão dos 30 próximos valores da série PETR4 usando Redes Neurais.

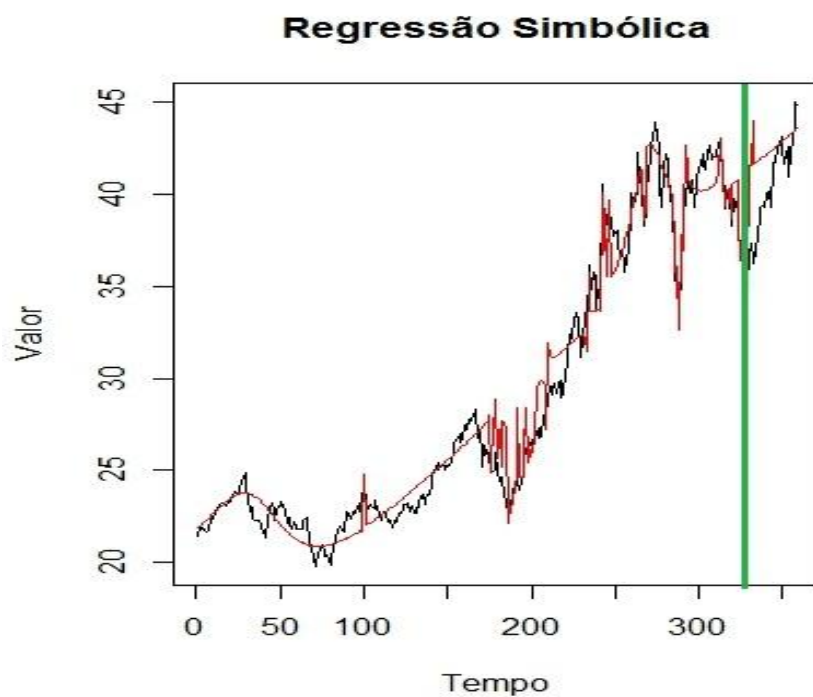


Figura 5.12: Previsão dos 30 próximos valores da série PETR4 usando Regressão Simbólica.

Assim como foi disponibilizado para os outros métodos, a Tabela 5.3 mostra alguns dos valores mais relevantes para cada uma das previsões feitas utilizando a série PETR4.

Tabela 5.3: Valores reais, valores previstos e erros nas previsões para a série PETR4

PETR4									
#	Real	Naïve		ARIMA		Rede Neural		Regressão Simbólica	
		Previsto	Erro Quadrático	Previsto	Erro Quadrático	Previsto	Erro Quadrático	Previsto	Erro Quadrático
1	36,76	34,91	3,422500	34,91	3,404834	35,22	2,383839	41,59	23,336314
2	37,35	34,91	5,953600	34,66	7,237500	35,52	3,361138	41,59	17,999429
3	36,46	34,91	2,402500	34,35	4,460278	35,81	0,422011	43,97	56,451424
4	36,21	34,91	1,690000	33,90	5,323860	36,10	0,013052	41,66	29,698873
5	36,90	34,91	3,960100	33,60	10,878942	36,37	0,279340	41,72	23,262000
6	37,78	34,91	8,236900	33,19	21,066538	36,64	1,307887	41,79	16,064201
7	38,79	34,91	15,054400	32,83	35,573410	36,89	3,612096	41,86	9,395414
8	39,25	34,91	18,835600	32,41	46,775848	37,13	4,494812	41,92	7,151051
9	39,40	34,91	20,160100	32,00	54,798477	37,36	4,173304	41,99	6,731287
10	39,25	34,91	18,835600	31,57	59,012476	37,57	2,819898	42,07	7,929236
11	40,00	34,91	25,908100	31,11	78,954978	37,77	4,970486	42,14	4,571873
12	40,02	34,91	26,112100	30,66	87,554854	37,96	4,257967	42,21	4,801522
13	40,16	34,91	27,562500	30,17	99,710367	38,13	4,125645	42,28	4,515249
14	39,28	34,91	19,096900	29,70	91,856158	38,29	0,984458	42,36	9,481062
15	41,48	34,91	43,164900	29,18	151,384698	38,43	9,278944	42,43	0,909807
16	41,97	34,91	49,843600	28,67	176,970122	38,57	11,576603	42,51	0,290494
17	41,70	34,91	46,104100	28,12	184,474160	38,69	9,063086	42,58	0,782345
18	42,55	34,91	58,369600	27,58	224,210952	38,80	14,059723	42,66	0,012185
19	42,65	34,91	59,907600	27,00	244,970822	38,90	14,056023	42,74	0,007499
20	43,10	34,91	67,076100	26,42	278,102982	38,99	16,878083	42,81	0,082306
21	42,51	34,91	57,760000	25,82	278,659942	39,07	11,808718	42,89	0,144328
22	41,60	34,91	44,756100	25,21	268,687228	39,15	6,015676	42,97	1,868586
23	42,15	34,91	52,417600	24,57	308,975282	39,21	8,623173	43,04	0,799711
24	42,59	34,91	58,982400	23,93	348,196262	39,27	11,004024	43,12	0,282812
25	40,90	34,91	35,880100	23,26	311,030473	39,33	2,478060	43,20	5,287948
26	42,20	34,91	53,144100	22,59	384,633125	39,37	7,990793	43,28	1,161036
27	43,00	34,91	65,448100	21,89	445,585778	39,42	12,848826	43,36	0,126500
28	43,85	34,91	79,923600	21,18	513,861456	39,45	19,332374	43,43	0,173048
29	45,00	34,91	101,808100	20,45	602,550264	39,49	30,396858	43,51	2,212574
30	44,85	34,91	98,803600	19,71	632,023103	39,52	28,446480	43,59	1,584544
		RMSE=	6,246654		14,096010		2,892884		2,811374

5.4.4 Resultados para VALE5

Seguindo o mesmo padrão de dados apresentados na sessão 5.4.3, as Figuras 5.13 a 5.16 apresentam as previsões feitas para a série VALE5.

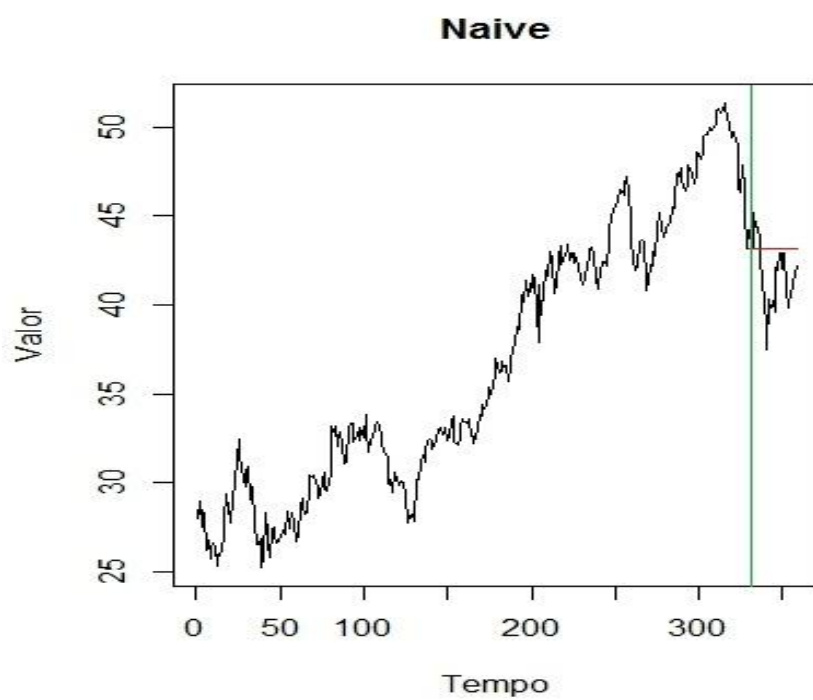


Figura 5.13: Previsão dos 30 próximos valores da série VALE5 usando Naive.

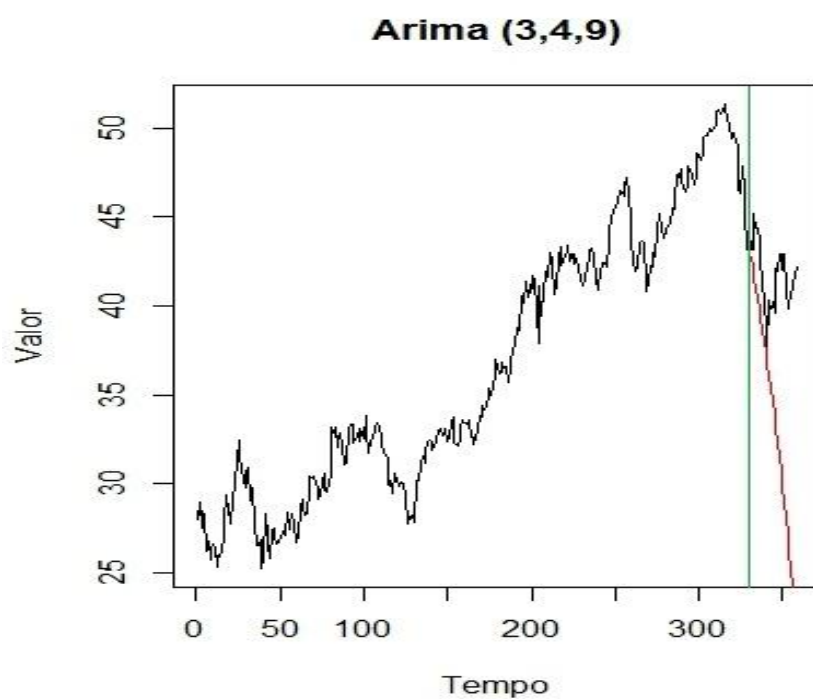


Figura 5.14: Previsão dos 30 próximos valores da série VALE5 usando ARIMA.

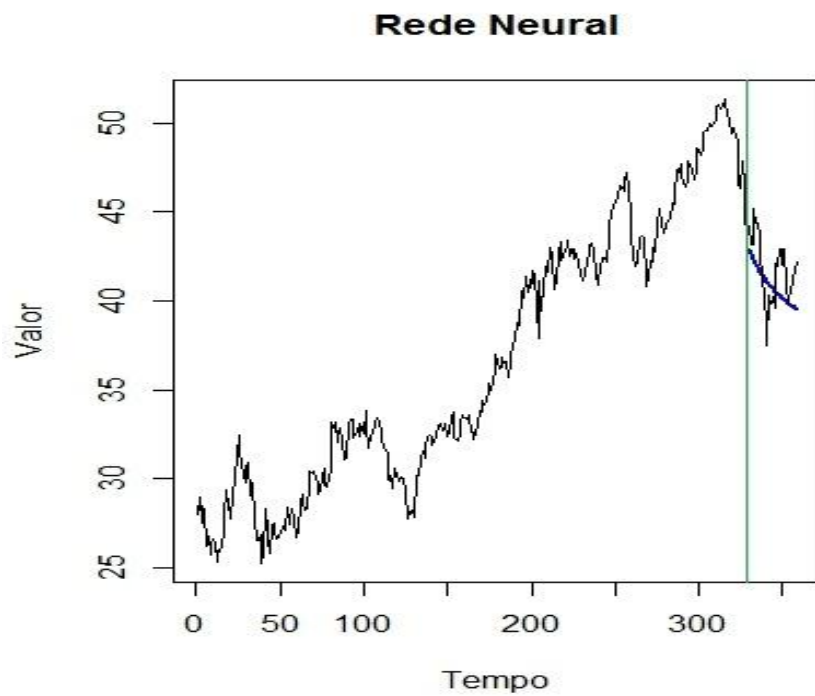


Figura 5.15: Previsão dos 30 próximos valores da série VALE5 usando Redes Neurais.

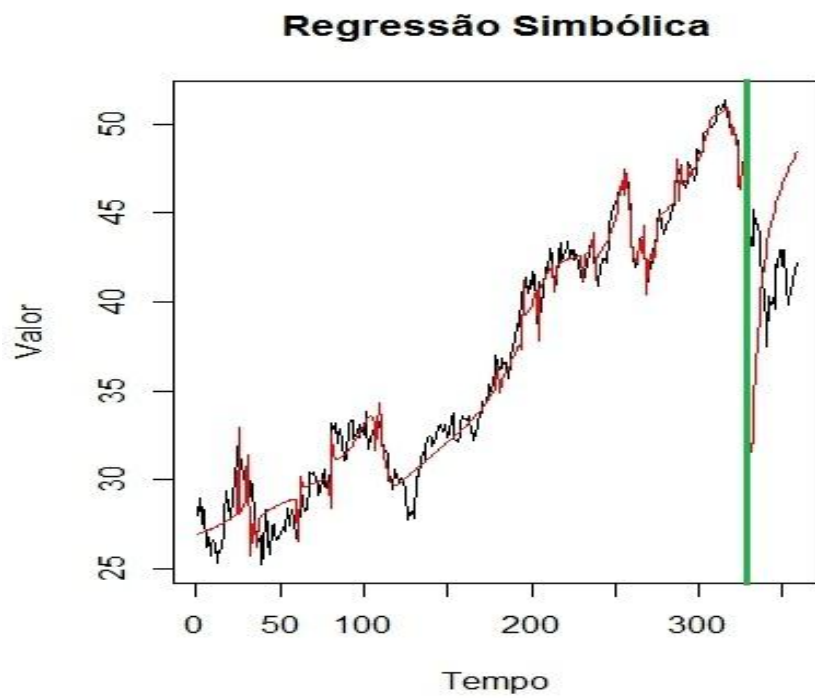


Figura 5.16: Previsão dos 30 próximos valores da série VALE5 usando Regressão Simbólica.

Ainda para este experimento, foram realizadas diversas execuções distintas para um mesmo conjunto de dados. O resultado escolhido para cada conjunto de entrada foi o que apresentou melhor fitness entre todas as execuções, ou seja, foi o que melhor aproximou os dados conhecidos das séries temporais.

A Tabela 5.4 mostra com maiores detalhes os valores das previsões realizadas e os valores dos erros obtidos através do experimento. Na próxima sessão será feito uma análise comparativa dos resultados obtidos com as diversas formas de previsão apresentadas neste trabalho.

Tabela 5.4: Valores reais, valores previstos e erros nas previsões para a série VALES5

VALES5									
#	Real	Naïve		ARIMA		Rede Neural		Regressão Simbólica	
		Previsto	Erro Quadrático	Previsto	Erro Quadrático	Previsto	Erro Quadrático	Previsto	Erro Quadrático
1	44,20	43,15	1,102500	42,86	1,799914	42,92	1,639902	32,69	132,381488
2	43,17	43,15	0,000400	42,75	0,175029	42,70	0,218725	31,58	134,235834
3	43,15	43,15	0,000000	42,27	0,773788	42,50	0,425693	32,36	116,422811
4	45,15	43,15	4,000000	41,54	13,029890	42,30	8,099285	34,40	115,521659
5	44,40	43,15	1,562500	41,01	11,510542	42,12	5,193849	36,28	65,961702
6	44,40	43,15	1,562500	40,41	15,896574	41,95	6,014678	37,88	42,515954
7	43,70	43,15	0,302500	39,94	14,156699	41,78	3,675208	39,21	20,134699
8	42,39	43,15	0,577600	39,29	9,626295	41,63	0,582834	40,32	4,266232
9	41,61	43,15	2,371600	38,66	8,689308	41,48	0,017454	41,26	0,122332
10	40,00	43,15	9,922500	38,07	3,736903	41,34	1,785884	42,06	4,231156
11	39,00	43,15	17,222500	37,40	2,571168	41,20	4,846809	42,74	14,013071
12	37,50	43,15	31,922500	36,75	0,556091	41,07	12,766299	43,34	34,123236
13	40,29	43,15	8,179600	36,06	17,906822	40,95	0,436037	43,87	12,803499
14	39,80	43,15	11,222500	35,36	19,740801	40,83	1,067505	44,34	20,580332
15	40,20	43,15	8,702500	34,64	30,877423	40,72	0,271737	44,76	20,764537
16	39,61	43,15	12,531600	33,89	32,711301	40,61	1,008586	45,14	30,548230
17	42,10	43,15	1,102500	33,13	80,393076	40,51	2,521977	45,48	11,449390
18	41,95	43,15	1,440000	32,34	92,276184	40,41	2,359414	45,80	14,837481
19	42,88	43,15	0,072900	31,54	128,697072	40,32	6,552804	46,10	10,342706
20	42,00	43,15	1,322500	30,71	127,553914	40,23	3,131867	46,37	19,091077
21	42,90	43,15	0,062500	29,85	170,370340	40,14	7,594602	46,62	13,873931
22	42,90	43,15	0,062500	28,97	194,046043	40,06	8,056506	46,86	15,718810
23	40,90	43,15	5,062500	28,06	164,766887	39,98	0,841966	47,09	38,329576
24	39,83	43,15	11,022400	27,13	161,201421	39,91	0,005843	47,31	55,884978
25	40,26	43,15	8,352100	26,18	198,314766	39,83	0,181876	47,51	52,558530
26	40,42	43,15	7,452900	25,19	231,854394	39,76	0,430939	47,70	53,065394
27	41,22	43,15	3,724900	24,18	290,250333	39,70	2,321556	47,89	44,505998
28	41,66	43,15	2,220100	23,14	342,838582	39,63	4,113650	48,07	41,096582
29	41,70	43,15	2,102500	22,08	385,069772	39,57	4,537852	48,24	42,817490
30	42,20	43,15	0,902500	20,98	450,267585	39,51	7,235069	48,41	38,570099
		RMSE=	2,280980		10,330632		1,806806		6,379051

5.5 Comparação e Análise dos Resultados

Considerando os resultados obtidos com os experimentos descritos nas seções anteriores, alguns detalhes ficam bastante claros. O primeiro ponto a ser destacado é que nenhuma das técnicas aqui analisadas foi capaz de prever com exatidão o comportamento das séries temporais devido, possivelmente, a sua natureza aleatória. O comportamento aleatório característico destas séries faz com que a tarefa de previsão seja extremamente difícil. Outro detalhe importante que é necessário frisar é que mesmo um algoritmo bastante utilizado como o ARIMA, por exemplo, é capaz de realizar uma predição muito boa para algumas séries, como foi o caso dos dados do Mercado Forex, e, ao mesmo tempo, fazer uma previsão bastante errônea, como aconteceu com os dados do Mercado de Ações.

Quando analisados caso a caso, percebemos alguns detalhes. No caso da série EURUSD, por exemplo, os resultados obtidos tanto pelo ARIMA quanto pelas Redes Neurais foram capazes de captar a tendência de alta do mercado. Por outro lado, o método Naïve gerou uma previsão mais conservadora em relação aos dados reais e a Regressão Simbólica apresentou um erro bastante elevado, embora seu resultado tenha sido de acordo com tanto a tendência de alta inicial quanto o início da tendência de baixa do final do período previsto.

Para a série GBPUSD, o modelo ARIMA também apresentou os melhores resultados, sendo este e as Redes Neurais, novamente capazes de manter certa conformidade com as tendências dos dados reais. O método Naïve tornou a realizar uma previsão bastante conservadora em relação ao que foi apresentado no período e, neste teste, foi ultrapassado em resultados pelo método da Regressão Simbólica. É importante ressaltar, também, que a Regressão Simbólica apresentou um erro ainda melhor do que o gerado pelas Redes Neurais, devido, possivelmente, a sua boa previsão na tendência de alta no início do período considerado.

Para as séries do Mercado de Ações, onde foram considerados mais pontos das séries originais, as previsões realizadas pela Regressão Simbólicas foram significativamente melhores. Na série PETR4, por exemplo, o método Naïve foi bastante conservador em sua previsão, mas foi melhor que o ARIMA, que previu uma queda nos valores, a qual não ocorreu. Nesta série, as Redes Neurais previram bem a tendência de alta, mas o melhor resultado foi o da Regressão Simbólica que conseguiu prever com menor taxa de erro essa mesma tendência.

Na última série, a VALE5, nenhum dos métodos apresentou um resultado que possa ser considerado satisfatório. O método Naïve, o qual era bastante conservador para tendências de alta, fez uma previsão muito otimista quando comparada aos dados reais. O ARIMA conseguiu prever a tendência de baixa inicial, mas acabou exagerando em suas previsões e não previu a reversão da série no período considerado. Para esse experimento, as Redes Neurais foram as melhores sendo capazes de acompanhar o movimento de descida do gráfico e suavizando sua curva quando a série apresentou a tendência de reversão. A Regressão Simbólica, por sua vez, não apresentou o pior dos resultados, sendo ela capaz de demonstrar, ainda que de uma forma um pouco extremista, todas as tendências vistas no gráfico para o período previsto.

Analisando os gráficos dos testes com Regressão Simbólica exclusivamente, verifica-se que quanto melhor for a aproximação dos dados conhecidos, melhor tende a ser a previsão dos valores futuros. Nos casos da série PETR4, por exemplo, a aproximação chegou a ser bastante precisa e a previsão, por sua vez, conseguiu seguir as

tendências de aumento e diminuição dos valores, embora extrapolando os valores em algumas ocasiões. Em alguns casos, uma aproximação satisfatória dos valores conhecidos da série em tempo de execução é obtida, contrastando com a previsão dos valores, que apresenta resultados muito distantes dos reais. Apesar disso, estes resultados estão de acordo com a tendência dos valores da série original.

A Tabela 5.5 reúne os valores referentes aos erros obtidos por cada técnica aplicada em cada um dos quatro grupos de dados utilizados nesse trabalho. Os resultados em verde e vermelho são os melhores e piores erros, respectivamente, considerando cada série analisada. Ainda na tabela, podemos perceber que a Regressão Simbólica ocupou posições bastante contrastantes em diferentes experimentos. Para alguns casos, esses resultados apresentados foram até superiores aos das Redes Neurais, consideradas tão boas ou melhores que o método ARIMA, como foi explicado na sessão 4.3. Ainda nesta tabela de dados, destaca-se a Regressão Simbólica como a melhor previsão feita para PETR4, além de ter feito uma previsão melhor que o ARIMA para a série VALE5.

Tabela 5.5: RMSE resultantes de todas as simulações

RMSE				
	Naïve	ARIMA	Redes Neurais	Regressão Simbólica
EURUSD	0,04182	0,01605	0,02896	0,12922
GBPUSD	0,03345	0,01605	0,02637	0,02617
PETR4	6,24665	14,09601	2,89288	2,81137
VALE5	2,28098	10,33063	1,80681	6,37905

Uma vantagem percebida na Regressão Simbólica foi justamente a capacidade de recuperar a função que descreve a aproximação da série considerada, como foi explicado na sessão 4.4. A obtenção dessa aproximação pode vir a ser útil se forem consideradas as possibilidades de realizações de diferentes testes nesta função.

Por fim, é importante perceber, tanto pela Tabela 5.5 quando pelos gráficos apresentados neste capítulo, que seja qual for o método de previsão, este apresenta grande dificuldade de prever séries com comportamentos tão caóticos quanto as séries pertencentes aos Mercados Financeiros. A maioria dos métodos apresentados obtiveram resultados considerados os melhores para alguns testes e resultados considerados os piores para outros testes. Dessa forma, é importante perceber que não existe um método que foi superior a todos em todos os testes e isso pode sugerir a não existência de um método perfeito, ao menos entre as técnicas utilizadas.

6 CONCLUSÕES E SUGESTÕES

Os Mercados Financeiros estão cheios de oportunidades para um investidor ou um *trader* que esteja disposto a se arriscar. Ainda assim, existem formas de controlar estes riscos através de técnicas de Análise Gráfica, também conhecida como Análise Técnica. Na prática, são utilizados conceitos extraídos e dependentes dos próprios valores gerados nestes mercados em um determinado tempo no passado, de forma a conseguir indícios de qual será o comportamento destes valores em um determinado tempo no futuro.

Aliada a Análise Técnica existem formas de aproximar uma série temporal descrita nestes mercados e tentar prever quais serão os próximos valores desta série. Dessa forma, o investidor dispõe de mais uma série de ferramentas para diminuir os riscos de seus investimentos pessoais. Neste trabalho foi apresentada a Regressão Simbólica de séries temporais utilizando Programação Genética, uma técnica relativamente recente nas áreas de pesquisa, com o objetivo de prever os próximos pontos destas séries.

Para constatar a eficácia da técnica, foi realizada uma série de experimentos comparativos baseando-se em algumas das séries mais utilizadas nos Mercados de Ações e Mercados Forex onde chegou-se a conclusão que a Regressão Simbólica pode apresentar previsões cujos erros são equivalentes ou até menores que algumas das outras técnicas consideradas mais eficientes hoje em dia. Sendo assim, é possível diminuir ainda mais os riscos dos investimentos incorporando este método às técnicas já utilizadas. Por outro lado, o esforço requerido para processar os dados de uma série utilizando este método faz com que exista a necessidade de explorar todo o paralelismo possível na Programação Genética, de forma a possibilitar o uso prático desta técnica em transações rápidas.

6.1 Sugestões para Trabalhos Futuros

Os resultados alcançados pela Regressão Simbólica quando comparados a métodos como o Naïve, o ARIMA e as Redes Neurais, sugeriram a existência de um potencial muito forte nessa técnica, mas que talvez ainda necessite de algum esforço para ser atingido. Portanto, visando desenvolver ainda mais este método tão promissor, ficam como sugestões para trabalhos futuros as seguintes alternativas:

- Utilização de maior volume de dados de forma que a Regressão obtenha mais informações sobre os padrões da série analisada.

- Implementação de Busca Local aliada a Regressão Simbólica de forma a aproximar com maior precisão a série temporal utilizada.
- Utilização do modelo de Ilhas apresentado em (KOZA, 1992) e implementação de um algoritmo que trabalhe em rede com outros computadores, podendo distribuir ainda mais o esforço de processamento.
- Em um uso específico para os mercados, alterar o modelo de previsão da série temporal para um modelo de decisão sobre quando o investidor deve comprar, vender ou realizar ação alguma com a mercadoria.

REFERÊNCIAS

ADLER, Joseph. **R in a Nutshell**. 2ª ed. [S.l.]: O'Reilly, 2012.

BÄCK, Thomas; SCHWEFEL, Hans-Paul. **Evolutionary Computation: Comments on the History and Current State**. IEEE Transactions on Evolutionary Computation, v. 1, n. 1, p. 3 – 17, Abr. 2007.

BANZHAF, Wolfgang; LEIER, Andre. **Evolution on Neutral Networks in Genetic Programming**. Genetic Programming Theory and Practice III, Springer, p. 207–222, 2006.

BECKER, Marcel Henrique. **Modelos para Previsão de Séries Temporais: uma Aplicação para a Taxa de Desemprego na Região Metropolitana de Porto Alegre**. 2010, 46 f. Trabalho de Graduação (Bacharelado em Estatística) – Instituto de Matemática, UFRGS, Porto Alegre.

BM&FBOVESPA: A Nova Bolsa. **Entenda o Mercado de Ações**. Disponível em: <<http://www.bmfbovespa.com.br/pt-br/educacional/iniciantes/mercado-de-acoes/entenda-o-mercado-de-acoes/entenda-o-mercado-de-acoes.aspx?idioma=pt-br>> Acesso em novembro/2012.

BM&FBOVESPA: A Nova Bolsa. **Sobre a Bolsa**. Disponível em: <<http://www.bmfbovespa.com.br/pt-br/intros/intro-sobre-a-bolsa.aspx?idioma=pt-br#>> Acesso em novembro/2012.

BM&FBOVESPA: A Nova Bolsa. **Índices**. Disponível em: <<http://www.bmfbovespa.com.br/indices/BuscarIndices.aspx?idioma=pt-br>> Acesso em novembro/2012.

BM&FBOVESPA: A Nova Bolsa. **Pregão Online BM&FBOVESPA**. Disponível em: <<http://pregao-online.bmfbovespa.com.br/Principal.aspx?idioma=pt-br#>>. Acesso em novembro/2012.

CARBONNEAU, Real; LAFRAMBOISE, Kevin; VAHIDOV, Rustav. **Applications of machine learning techniques for supply chain demand forecasting**. European Journal of Operational Research 184, p. 1140 – 1154, 2008.

CVM. **Mercado Forex: Série Alertas**. Disponível em: <<http://www.cvm.gov.br/port/taxas/mercadoForex.pdf>>. Acesso em outubro/2012.

DAVIDSON, Alexander. **How the Global Financial Markets Really Work: The Definite Guide to Understanding International Investment and Money Flows**. 1^a ed. Philadelphia: Kogan Page Limited, 2009.

FERNANDES, Fabrício. **Integração de métodos quantitativos e qualitativos para previsão de demanda no setor de autopeças**. 2010, 27 f. Trabalho de Graduação (Curso de Engenharia de Produção) – Escola de Engenharia, UFRGS, Porto Alegre. Disponível em: <<http://hdl.handle.net/10183/32218>>. Acesso em Dezembro/2012.

FRANK, R. J.; DAVEY N.; HUNT, S. P. **Time Series Prediction and Neural Networks**. Journal of Intelligent and Robotic Systems, v. 31, p. 91 – 103, 2001.

FX HISTORICAL DATA. Disponível em: <<http://www.fxhistoricaldata.com>>. Acesso em: novembro/2012.

GHEYAS Iffat A.; SMITH, Leslie S. **A Neural Network Approach to Time Series Forecasting**. Proceedings of the World Congress on Engineering, v. 2, p. 1292, 2009.

GODOY, José; MEDINA, Luiz Gustavo; JUNIOR, Marco Antônio Gazel. **Investindo em ações: os primeiros passos: as dicas do Sr. Alceu**. 3^a ed. São Paulo: Saraiva, 2006.

GRINGS, Alexandre. **Regressão Simbólica via Programação Genética: um Estudo de Caso com Modelagem Geofísica**. 2006, 133 f. Dissertação (Mestrado em Ciência da Computação) – Faculdade de Computação, Universidade Federal de Uberlândia, Uberlândia.

HOANG, Tuan Hao et al. **The Importance of Local Search: A Grammar Based Approach to Environmental Time Series Modeling**. Genetic Programming Theory and Practice III, Springer, p. 159–175, 2006.

HORTON, Nicholas J., KLEINMAN, Ken. **Using R for Data Management, Statistical Analysis, and Graphics**. 1^a ed. [S.l.]: CRC Press, 2011.

HYNDMAN, Rob J.; KHANDAKAR, Yeasmin. **Automatic Data Series Forecasting: The forecast Package for R**. Journal of Statistical Software, v. 27, n. 3, 2008.

JONES, Lee W.; AL-SHAKRAN, Sameer H.; KOZA, John R. **Automated Design of a Previously Patented Aspherical Optical Lens System by Means of Genetic Programming**. Genetic Programming Theory and Practice III, Springer, p. 33–48, 2006.

KLÚČIK, Miroslav; JURIOVÁ, Jana; KLÚČIK, Marian. **Time Series Modeling with Genetic Programming Relative to ARIMA Models**. Conferences on New Techniques and Technologies for Statistics. p. 17-27. Brussels: Eurostat, 2009.

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

LEICHT, Gordon; TANNER, J. Ernest. **Economic Forecast Evaluation: Profits Versus the Conventional Error Measures**. The American Economic Review, v. 18, n. 3, p. 580 – 590, 1991.

LEVINSON, Mark. **Guide to Financial Markets**. 4ª ed. Londres: The Economist, 2005.

LI, Jin. **FGP: A Genetic Programming Based Financial Forecasting Tool**. 2000. 190f. Tese (Doutor de Filosofia em Ciência da Computação) – Departamento de Ciência da Computação, Universidade de Essex, Colchester.

LUKE, Sean et al. **ECJ 20 - A Java-based Evolutionary Computation Research System**. Disponível em: <<http://cs.gmu.edu/~eclab/projects/ecj/>>. Acesso em outubro/2012.

LUKE, Sean. **The ECJ Owner's Manual**. 0ª ed. [S.l.: s.n], 2010. Disponível em: <<http://cs.gmu.edu/~eclab/projects/ecj/docs/manual/manual.pdf>>. Acesso em dezembro/2012.

LUKE, Sean. **Essentials of Metaheuristics**. 1ª ed. [S.l.]: Lulu, 2012. Disponível em: <<http://cs.gmu.edu/~sean/book/metaheuristics/Essentials.pdf>>. Acesso em dezembro/2012.

MATSURA, Eduardo. **Comprar ou Vender? Como Investir na Bolsa Utilizando Análise Gráfica**. 6ª ed. São Paulo: Saraiva, 2007.

MURPHY, John J. **Technical Analysis of the Financial Markets**. 1ª ed. New York: New York Institute of Finance, 1999.

NIST/SEMATECH. **e-Handbook of Statistical Methods**. Disponível em: <<http://www.itl.nist.gov/div898/handbook/>>. Acesso em outubro/2012.

PONSI, Ed. **Forex Patterns & Probabilities: trading strategies for trending & range-bound markets**. 1ª ed. Hoboken: John Wiley & Sons, Inc., 2007.

PRING, Martin J. **Technical Analysis Explained: the Successful investor's Guide to Spotting Investment Trends and Turning Points**. 4ª ed. New York: 2002.

RATNIEKS, Ianes. **Métodos de Previsão: Aplicação da Metodologia de Box e Jenkins ao Varejo Brasileiro – O Caso das Lojas Americanas S.A.** 2010, 90 f. Trabalho de Graduação (Bacharelado em Economia) – Faculdade de Ciências Econômicas, UFRGS, Porto Alegre.

ROCKEFELLER, Barbara. **Technical Analysis for Dummies**. 2ª ed. Indianapolis: Wiley Publishing, Inc., 2011.

RPROJECT. **The R Project for Statistical Computing**. Disponível em: <<http://www.r-project.org/>>. Acesso em: dezembro/2012.

RSTUDIO. **RStudio – Home**. Disponível em: <<http://www.rstudio.com/>>. Acesso em: dezembro/2012.

SESHADRI, Mukund. **Comprehensibility, Overfitting and Co-Evolution in Genetic Programming for Technical Trading Rules**. 2003. 87f. Tese (Mestrado em Ciência da Computação) – Worcester Polytechnic Institute.

SHABRI, Ani bin. **Comparison of Time Series Forecasting Methods Using Neural Networks and Box-Jenkins Models**. *Mathematika* 17: p. 25–32, 2001.

SIM, Kwang Mong; WANG, Yuanshi. **Evolutionary Asymmetric Games for Modeling Systems of Partially Cooperative Agents**. *IEEE Transactions On Evolutionary Computation*, v. 9, n. 6, IEEE, p. 603-614, 2005.

SPECTOR, Lee; KLEIN, Jon. **Trivial Geography in Genetic Programming**. Genetic Programming Theory and Practice III, Springer, p. 109–124, 2006.

TEETOR, Paul. **25 Recipes for Getting Started with R**. 1ª ed. [S.l.]: O'Reilly, 2011.

TUCI, Elio; MASSERA, Gianluca; NOLFI, Stefano. **Active Categorical Perception of Object Shapes in a Simulated Anthropomorphic Robotic Arm**. IEEE Transactions On Evolutionary Computation, v. 14, n. 6, IEEE, p. 885-899, 2010.

UOL ECONOMIA. **Cotações - Bolsas de Valores**. Disponível em: <<http://economia.uol.com.br/cotacoes/bolsas/>>. Acesso em: novembro/2012.

WEI, William W. S. **Time Series Analysis: univariable and multivariable methods**. 2ª ed. Boston: Addison-Wesley, 2006.

WURGLER, Jeffrey. **Financial markets and the allocation of capital**. Journal of Financial Economics. n. 58, p. 187 – 214, 2000.

ZHANG, Yongqiang; CHENG, Huifang; YUAN, Ruilan. **Improved Genetic Programming Algorithm Applied to Symbolic Regression and Software Reliability Modeling**. Journal of Software Engineering and Applications, v. 2, n. 5, p. 354 - 360, Dez. 2009. Disponível em: <<http://www.scirp.org/Journal/Home.aspx?IssueID=185>>. Acesso em outubro/2012.

ZULUTRADE. **Gráficos de Taxas**. Disponível em: <http://v4.zulutrade.com/charts>>. Acesso em novembro/2012.

GLOSSÁRIO

Crossover: Métodos de cruzamento entre dois indivíduos na Programação Genética.

Fitness: Valor que determina o quão próximo de uma solução um determinado indivíduo está.

Forex: Mercado de Divisas ou Mercado Cambial.

Método Full: Método de criação de árvores na Programação Genética.

Método Grow: Método de criação de árvores na Programação Genética.

Perceptron Multicamada: Padrão para Redes Neurais que utilizam múltiplos neurônios (Perceptrons) conectados em três camadas (entrada, oculta e saída).

Ramped Half-and-Half: Métodos de criação de árvores na Programação Genética, baseado nos métodos *Full* e *Grow*.

Rank Method: Método de seleção na Programação Genética baseado na posição de um indivíduo.

Trader: Investidor de curtíssimo prazo. Busca lucros em operações iniciadas e finalizadas no mesmo dia.

Tournament Method: Método de seleção na Programação Genética baseado na competição entre um subgrupo de indivíduos.