**CENTRO UNIVERSITÁRIO FEEVALE**

**FERNANDO RAFAEL STAHNKE**

# USO DE *DATA MINING* NO MERCADO FINANCEIRO

**Novo Hamburgo, Dezembro de 2008.**

**FERNANDO RAFAEL STAHNKE**

# USO DE *DATA MINING* NO MERCADO FINANCEIRO

**Centro Universitário Feevale**
**Instituto de Ciências Exatas e Tecnológicas**
**Curso de Ciência da Computação**
**Trabalho de Conclusão de Curso**

**Professor Orientador: Juliano Varella de Carvalho**

**Novo Hamburgo, Dezembro de 2008.**

**AGRADECIMENTOS**

Gostaria de agradecer a todos os que, de alguma maneira, contribuíram para a realização desse trabalho de conclusão, em especial:

Ao professor orientador Juliano Varella de Carvalho e ao amigo Felipe Noer, pela disposição, dedicação e prestabilidade.

**RESUMO**

Enriquecer com investimento em bolsas de valores é o sonho de muitos investidores. Mas, comprar e vender ações no momento certo requer cautela e informações confiáveis. Hoje, devido à facilidade do acesso a estas negociações através de sistemas como o *home-broker*, a participação de investidores considerados pessoa física na Bovespa (Bolsa de Valores de São Paulo) é cada vez mais significativa. Conforme pesquisas, a participação destes investidores correspondeu a 25% do giro financeiro total da Bovespa no mês de fevereiro de 2008, com uma movimentação total de R$ 116,58 bilhões de reais. Segundo autores consultados, a crescente complexidade dos instrumentos de negociação do mercado financeiro, assim como o acesso a novas tecnologias de processamento da informação, estimulam o desenvolvimento de novos sistemas de análise e operação; inclusive com o uso de técnicas de Inteligência Artificial. O uso de técnicas de mineração de dados e inteligência artificial no mercado financeiro para determinação de estratégias e prever as tendências de alta e baixa já é realidade desde 1986 nos Estados Unidos. Primeiro, através da aplicação de sistemas baseados em regras heurísticas, com o uso de sistemas especialistas. Após, com o uso de análise estatística e, finalmente, com o uso de redes neurais artificiais. Todas as opções têm vantagens e desvantagens. Logo, este trabalho tem como objetivo discutir e propor o uso de técnicas de mineração de dados para a identificação de padrões de comportamentos hoje despercebidos pelos investidores e, com isso, determinar a tendência futura dos ativos do mercado à vista. Com base em pesquisas bibliográficas, serão aplicadas as tecnologias de redes neurais e árvores de decisão, comparando seus resultados. Após a obtenção de padrões, pretende-se aplicar o conhecimento adquirido no mercado acionário brasileiro, onde será determinada a validade ou não da metodologia proposta, seguindo os padrões identificados na pesquisa.

**Palavras-chave:** Ação. Mercado à vista. Mineração de Dados. Descoberta de Conhecimento. Padrões.

**ABSTRACT**

Getting rich with investments in the stock market is a dream for many investors. However, buying and selling shares at the right moment requires caution and reliable information. Nowadays, due to the facility to access these negotiations through systems such as home-broker, the participation of investors considered individuals at Bovespa ( Stock Market in São Paulo) is more and more significant. According to researches, the participation of these investors amounts to 25% of the total financial spin at Bovespa in February 2008, with a total handling of R$ 116,58 billions Reals. According to authors consulted, the increasing complexity of the instruments of negotiations of the financial market, as well as the access to new information processing technologies, stimulate the development of new analysis and operation systems; including the use of Artificial Intelligence techniques. The use of data mining and artificial intelligence in the financial market to determine the strategies and foresee the tendencies of high and low is already reality in the United States of America since 1986. First, through the application of systems based on heuristics rules, with the use of systems specialists. After that, with the use of statistical analysis and, finally, with the use of artificial neural networks. There are advantages and disadvantages in all of the options. Therefore, this paper aims to discuss and propose the use of data mining techniques to identify behavior patterns which are unnoticed nowadays by investors and, thereby, to determine the future tendency of the cash market. Based on bibliographic research, the technologies of neural networks and decision trees will be applied, comparing the results. After obtaining the patterns, there is the intention to apply the knowledge acquired in cash market at Bovespa. In this environment, there will be determined the validity or not of the methodology proposed by the simulation of negotiations of shares, following the patterns identified in the research.

**Key words:** Cash. Cash market. Data Mining. Discovery of knowledge. Standards.

**LISTA DE TABELAS**

## LISTA DE GRÁFICOS

## LISTA DE FIGURAS

## LISTA DE QUADROS

## LISTA DE ANEXOS

**SUMÁRIO**

# INTRODUÇÃO

Atualmente, com o intenso intercâmbio entre os países, cada vez mais o mercado acionário vem adquirindo uma crescente importância no cenário financeiro de todo o mundo. Por ser um importante canal na captação de recursos, o mercado acionário também se constitui em uma importante opção de investimento para pessoas e instituições, fortemente amparado pela evolução tecnológica e facilidade de acesso de qualquer pessoa a este ambiente.

Enriquecer com investimento em bolsas de valores é o sonho de muitos investidores. Mas, comprar e vender ações no momento certo requer cautela e informações confiáveis. Hoje, devido à facilidade do acesso a estas negociações através de sistemas como o *home-broker*, a participação de investidores considerados pessoa física na Bovespa (Bolsa de Valores de São Paulo) é cada vez mais significativa. Conforme pesquisas, a participação destes investidores correspondeu a 25% do giro financeiro total da Bovespa no mês de fevereiro de 2008, com uma movimentação total de R$ 116,58 bilhões de reais (FOLHA ONLINE, 2008).

Atualmente, em todos os setores da sociedade o volume de dados armazenados é gigantesco e continua a crescer rapidamente a todo instante. Infelizmente, o ser humano não é capaz de interpretar tamanha quantidade de dados e, possivelmente, muitas informações e conhecimento estão sendo desperdiçados, ficando ocultos nas bases de dados utilizadas (REZENDE, 2005). Em conseqüência disto, há a necessidade de desenvolver novas ferramentas, utilizando técnicas de extração de conhecimento (CARVALHO, 2000).

Além disso, a sociedade na era do conhecimento está impondo uma competitividade cada vez maior entre as empresas, forçando-as a inovar e adquirir conhecimento de forma sucessiva. (REZENDE, 2005). Este fato não é diferente em ambientes como a bolsa de valores, onde a negociação entre cada investidor e corretora presente faz o “giro” de grandes montantes financeiros. Desta forma, a informação adicional, correta e precisa certamente é um fator de sucesso ou fracasso entre os investidores.

O uso de *Data Mining* neste contexto, mais especificamente no mercado à vista, visa definir uma nova forma de auxiliar o investidor na chamada “Análise Técnica”, que atualmente é executada através do estudo dos gráficos formados pelas cotações dos preços das ações das empresas de capital aberto. A aplicação de técnicas de mineração de dados baseia-se no fato de que estes gráficos são construídos com base nos preços das ações em um período de tempo.

Logo, este estudo tem como objetivo o uso de técnicas de mineração de dados para a identificação de padrões de comportamentos hoje despercebidos pelos investidores e, com isso, determinar a tendência futura dos ativos do mercado à vista.

# 1 O MERCADO FINANCEIRO

O conhecimento do mercado financeiro brasileiro tem como objetivo esclarecer a função e localização das bolsas de valores dentro desta estrutura.

Sistema financeiro é o conjunto de instituições e instrumentos envolvidos com o fluxo de recursos monetários entre os agentes econômicos e a regularização deste processo (BOVESPA, 2007[1]). O Conselho Monetário Nacional - CMN, seu organismo maior, presidido pelo ministro da Fazenda, é quem define as diretrizes de atuação do sistema. Este sistema permite a transferência de recursos dos ofertadores finais, que se encontram em *superávit* financeiro para os tomadores finais, em *déficit* financeiro. Este fluxo cria condições onde os títulos e valores mobiliários tenham liquidez no mercado. As instituições financeiras deste sistema podem ser classificadas segundo a natureza de suas obrigações, distinguindo-as em bancárias e não bancárias e também pelos tipos de operação que estão autorizadas a realizar, classificando-as em instituições de crédito, distribuidoras de títulos e valores mobiliários (CAVALCANTE; MISUMI; RUDGE, 2005).

Basicamente, o sistema financeiro nacional é constituído por um subsistema normativo e um operativo. O subsistema normativo regula e controla o sistema operativo através de normas legais, emitidas pela autoridade monetária ou pelos agentes financeiros do governo. O subsistema operativo é constituído por instituições financeiras públicas ou privadas que atuam no mercado financeiro.

Como o objetivo deste trabalho é a exploração do mercado de ações da Bovespa (Bolsa de Valores de São Paulo), sendo esta uma instituição distribuidora de títulos e valores

mobiliários e, entidade operativa do sistema financeiro brasileiro, foi focado o estudo nesta instituição e nas entidades diretamente relacionadas a ela, como corretoras e agentes autônomos de investimento. Estas entidades formam o chamado mercado de capitais.

A entidade normativa deste sistema é representada pela Comissão de Valores Mobiliários (CVM), sendo uma entidade autárquica, vinculada ao Ministério da Fazenda (CAVALCANTE; MISUMI; RUDGE, 2005). A CVM é voltada para o desenvolvimento, disciplina e fiscalização do mercado de valores mobiliários não emitidos pelo sistema financeiro e pelo Tesouro Nacional, (FORTUNA, 1997), configurando basicamente o mercado de ações e debêntures.

São objetivos da CVM assegurar o funcionamento eficiente e regular os mercados de bolsa e instituições auxiliares que operam nestes mercados, regulamentar, orientar e fiscalizar fundos de investimentos, estimular a aplicação de poupança no mercado acionário. Além disso, a CVM deve proteger os titulares de valores mobiliários contra atos ilegais e emissões irregulares que tenham a finalidade de manipular os mercados primários e secundários de ações e fiscalizar a emissão, o registro, a distribuição e a negociação de títulos emitidos pelas entidades anônimas de capital aberto. Desta forma, o objetivo final da CVM é o fortalecimento do mercado de ações, abrangendo todas as matérias referentes ao mercado de valores mobiliàrios.

## 1.1 Mercado de Capitais

O mercado de capitais é um sistema de distribuição de valores mobiliários composto pela bolsa de valores, sociedades corretoras e outras instituições financeiras autorizadas (BOVESPA, 2007[1]) tem como objetivo proporcionar liquidez aos títulos emitidos por empresas de capital aberto e viabilizar seu processo de capitalização. Estes títulos podem ser ações, debêntures conversíveis em ações, bônus de subscrição e *commercial papers*.

O mercado de ações, conforme Fortuna (1997) é dividido em duas partes. O **mercado primário**, composto pelas instituições bancárias em operações de subscrição

(*Underwriting*) envolvendo ações, debêntures e outros títulos e o **mercado secundário**, onde as ações são comercializadas através de bolsas de valores.

### 1.1.1 Ações

Ações são títulos de renda variável emitidos por Sociedades Anônimas que representam a menor fração do capital social da empresa, tornando o possuidor deste título dono de uma fração do capital social da empresa que o emitiu (BOVESPA, 2007[1]). Normalmente, não possuem prazo de resgate e podem ser negociadas em mercados organizados como as bolsas de valores. Conforme Cavalcante, Misumi e Rudge (2005), por serem objeto de negociações diárias, o preço das ações variam de acordo com o interesse dos investidores. As ações são conversíveis em dinheiro, a qualquer momento, pela negociação em bolsas de valores ou mercado de balcão.

No Brasil, existem ações dos tipos ordinárias e preferenciais. As **Ações Ordinárias (ON)** são aquelas que proporcionam ao acionista direito de voto em assembléias, ou seja, determinam o destino da empresa e conferem a ele participação nos resultados da empresa a qual a ação se refere. Já as **Ações Preferenciais (PN)** são ações onde o acionista tem a prioridade no recebimento de dividendos e no reembolso de capital em caso de dissolução da sociedade, sem direito a voto (BOVESPA, 2007[1]).

## 1.2 Bolsas de Valores

Constituem-se como associações civis, sem fins lucrativos ou sociedades anônimas e têm como objetivo manter um local adequado ao encontro de seus membros e a realizações, entre eles, de transações de compra e venda de títulos e valores mobiliários, em mercado livre e aberto, organizado e fiscalizado por seus membros, pela autoridade monetária e pela CVM. (CAVALCANTE, MISUMI e RUDGE,2005). As bolsas propiciam liquidez às aplicações, fornecendo um preço de referência para os ativos negociados por intermédio de um mercado contínuo, representado por seus pregões diários.

No Brasil atualmente, existe em operação a BM&FBOVESPA S.A. - Bolsa de Valores, Mercadorias e Futuros, criada em 2008 com a integração entre Bolsa de Mercadorias & Futuros (BM&F) e Bolsa de Valores de São Paulo (BOVESPA), formando assim, a terceira maior bolsa do mundo em valor de mercado, a segunda das Américas e a líder no continente latino-americano (BOVESPA[5], 2008). A bovespa será o ambiente abordado por este estudo.

### 1.2.1 BM&FBOVESPA S.A

A antiga Bovespa (Bolsa de Valores de São Paulo) era o maior centro de negociações com ações da América Latina, fundada em 1890 e, desde então, incorporando toda a evolução política e econômica brasileira (CAVALCANTE; MISUMI; RUDGE, 2005). Éra a entidade responsável por diversas iniciativas visando a qualidade dos serviços prestados como o *Mega Bolsa*, *Home Broke* e *After-Market*. Em 2001, visando fortalecer o mercado brasileiro para a crescente globalização dos negócios, todas as bolsas regionais brasileiras foram integradas em um só mercado, o da Bovespa. Em 2008 a Bovespa foi integrada com a BM&F (Bolsa de Mercadorias e Futuros). Este estudo foi desenvolvido com estas instituições ainda separadas.

A Bovespa, como todas as bolsas de valores, segue os mesmos objetivos e conceitos legais e sociais citados anteriormente. Possui também o poder de auto-regulação, permitindo-a criar e fiscalizar procedimentos e normas para os agentes que nela atuam. Os principais títulos negociados na Bovespa além das ações são debêntures, *commercial Paper,* opções de compra e venda de ações, quotas de fundos, bônus de subscrição, recibos de carteiras de ações, títulos públicos e certificados de depósitos de ações.

Em 2007, a Bovespa negociou R$ 1,2 trilhão de reais, e uma média diária de R$ 4,9 bilhões de reais em valor financeiro e 153 mil negócios. Conforme dados da instituição, o valor de mercado das 404 empresas com ações negociadas na Bovespa somou R$ 2,5 trilhões em dezembro de 2007, crescimento de 60,4% em relação ao ano anterior, quando a capitalização das companhias na bolsa chegou a R$ 1,54 trilhão (BOVESPA, 2007[2]).

A participação dos tipos de investidores em cada um dos mercados da Bovespa é indicada conforme tabela a seguir, onde o mercado à vista se destaca em relação aos outros mercados oferecidos pela instituição.

**TABELA 1.1** – PARTICIPAÇÃO DOS INVESTIDORES NA BOVESPA

**Participação dos Investidores na Bovespa ( Compras + Vendas) Fevereiro/2008**

| Tipos de Investidores | Vista (R$) | Termo (R$) | Opções (R$) | Exerc. De Opções (R$) | Outros (R$) | Total (R$) | (%) |
|---|---|---|---|---|---|---|---|
| Pessoas Físicas | 50.904.478.631 | 2.412.116.234 | 4.531.863.489 | 1.285.810.028 | 63.337.671 | 59.197.606.053 | 25,48 |
| Institucionais | 59.182.002.539 | 2.756.511.599 | 1.651.645.438 | 3.570.486.962 | 98.291.328 | 67.258.937.866 | 28,95 |
| Investidores Estrangeiros | 77.971.324.289 | 512.771.581 | 376.510.686 | 1.653.933.416 | 12.957.296 | 80.527.497.268 | 34,66 |
| Empresas Públicas e Privadas | 3.417.436.437 | 410.596.237 | 173.335.537 | 349.765.064 | 166.364.335 | 4.517.497.610 | 1,94 |
| Instituições Financeiras | 16.023.473.914 | 2.114.106.149 | 582.908.863 | 1.812.748.171 | 1.439.823 | 20.534.676.920 | 8,84 |
| Outros | 263.478.129 | 9.032.430 | 7.030.529 | 2.325.664 | 96.592 | 281.963.344 | 0,12 |
| **Total Geral** | **207.762.193.939** | **8.215.134.230** | **7.323.294.542** | **8.675.069.305** | **342.487.045** | **232.318.179.061** | |

**Fonte:** BOVESPA, 2007[3], p12.

O volume total de negociações da Bovespa em 2007 mostrou significativa participação dos investidores tipo pessoa física, com cerca de 22% do volume total negociado do mês de agosto, conforme mostrado no gráfico 1.1. O percentual de participação deste tipo de investidor no mercado de ações é incentivado principalmente pelas inovações que tem como objetivo a facilitação do acesso ao mercado, como o *Home Broker*. O *Home Broker*, ferramenta da internet oferecida aos investidores, atingiu em dezembro de 2007 um volume total negociado de R$ 19,5 milhões, com uma média de mais de 121 mil negócios por dia, representando 31% no número total de negócios. (BOVESPA, 2007[3]).

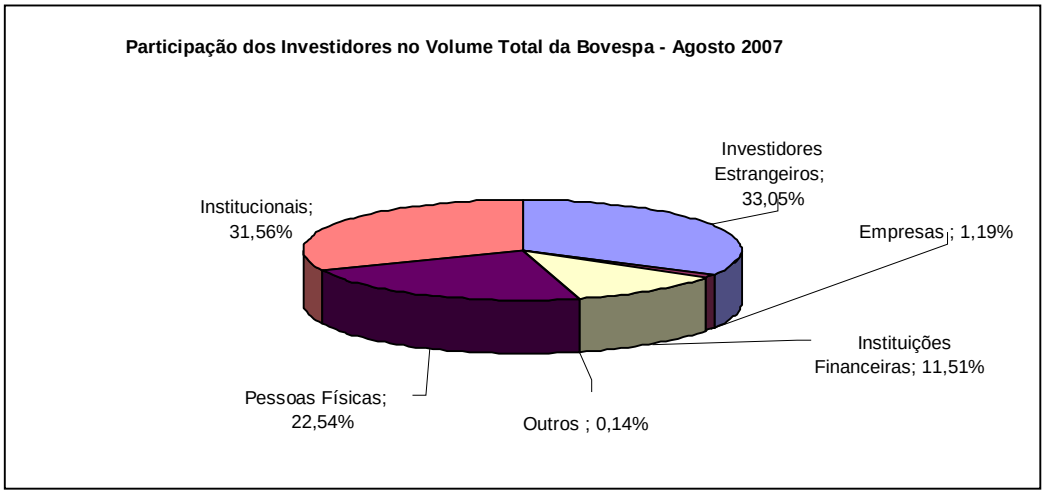


**GRÁFICO 1.1** - PARTICIPAÇÃO DOS INVESTIDORES NO VOLUME TOTAL DA BOVESPA - AGOSTO 2007.
**Fonte:** BOVESPA, 2007[4], p1.

Os mercados disponíveis para negociações na Bovespa são os mercados à vista, a termo e de opções, onde as negociações ocorrem exclusivamente no mercado eletrônico. Neste estudo são abordados os dados referentes ao mercado à vista, sendo em 2007 o mercado responsável por 88,7% das negociações da Bovespa e somando R$ 103,4 bilhões. Este mercado representa operações à vista de compra ou venda de títulos em lotes padrões de 100, 1000, 10.000 ou 100.000 ações e em lotes fracionários. Caracteriza-se pela entrega dos títulos vendidos no 2° dia útil após a realização do negócio (liquidação física) e pela liquidação financeira (pagamento e recebimento) no 3° dia útil após a negociação dos títulos, processada pela Companhia Brasileira de Liquidação e Custódia – CBLC (BOVESPA, 2007[1]).

### 1.3 CBLC

Essa é a empresa responsável pela compensação, liquidação e controle de risco das operações realizadas na Bovespa, assumindo o serviço de custódia fungível[1] que anteriormente era executado pela Bovespa. Esta empresa atua nos mercados à vista e de liquidação futura, tem a incumbência do registro e controle das operações de empréstimo de títulos, sendo hoje, responsável pela guarda de mais de 6 trilhões de ações de companhias

[1] Custódia em que os valores depositados podem não ser os mesmos quando de sua saída, embora sejam da mesma espécie, qualidade e quantidade.

abertas, certificados de investimentos e debêntures. A finalidade de sua criação foi propiciar ao mercado financeiro brasileiro uma estrutura moderna e eficiente para a realização de atividades de compensação, liquidação, custódia e controle de risco (BOVESPA, 2007[1]).

### 1.4 Sociedades Corretoras

A principal função destas instituições é promover de forma eficiente a aproximação entre compradores e vendedores de títulos e valores monetários, permitindo a negociação adequada através de operações nas bolsas de valores e mercadorias. Desta forma, unifica o mercado, fornecendo segurança ao sistema e liquidez aos títulos transacionados.

Estas instituições operam exclusivamente na bolsa de valores na qual são membros, com títulos e valores mobiliários de negociação autorizada no mercado aberto e em operações de câmbio. A operação consiste em comprar, vender e distribuir os títulos e valores mobiliários, por conta de terceiros e efetuar lançamentos públicos de ações, além de administrar carteiras de valores e custodiar títulos e valores mobiliários. Podem ainda executar tarefas como instituir, organizar e administrar fundos e clubes de investimentos, prestar serviços como transferência de títulos, desdobramento de cautelas, recebimento de juros, dividendos ou encarregar-se da subscrição de títulos de valores mobiliários.

Os investidores brasileiros podem participar do mercado de capitais, de duas formas: coletiva ou individualmente. De forma coletiva, através da aquisição de cotas de clubes de investimentos ou fundos mútuos de ações. **Fundos de investimentos** são entidades que, pela emissão de títulos de investimento próprios, concentram capitais de diversos investidores para aplicação em carteiras diversificadas de títulos, valores mobiliários, instrumentos financeiros, derivados ou *commodities* negociadas em bolsas de mercadorias e futuros (CAVALCANTE; MISUMI; RUDGE, 2005). **Clubes de investimentos** são condomínios constituídos por pessoas físicas para aplicação de recursos comuns em títulos e valores mobiliários. Embora represente uma pessoa jurídica, é tratado nas estatísticas da Bovespa como movimentações de pessoas físicas.

De forma individual, o investidor pode participar através de uma sociedade corretora demonstrando suas intenções de compra ou venda através do lançamento de ordens diretamente no mercado de capitais.

### 1.5 Análise de Ações

Atualmente, de acordo com os princípios básicos existem dois tipos de análises de investimentos utilizados pelos investidores individuais, sendo a análise fundamentalista e a análise técnica.

#### 1.5.1 Análise Fundamentalista

A análise fundamentalista baseia-se nos fatores econômicos e fundamentais das empresas, como ramo de atuação, perspectivas de mercado, imagem, além das análises dos demonstrativos financeiros e de relatórios de administração emitidos por elas, sendo utilizada em aplicações de longo prazo (MATSURA, 2007).

Para estes investidores, os principais fatores que determinam o preço justo de uma ação são: o custo de oportunidade de renda fixa; prêmio de risco inerente ao investimento da empresa; liquidação da ação no mercado secundário, geração de resultados e taxa de crescimento (CAVALCANTE; MISUMI; RUDGE, 2005). Para acompanhar estes fatores, estes analistas utilizam diversos indicadores, como por exemplo: Preço/Lucro (P/L), Preço/Valor Patrimonial, Preço/Terminais Telefônicos, Preço/Vendas líquidas, Vendas/Metro quadrado de loja, Preço/Geração Caixa Operacional/ Preço/Ebtida[2] e Enterprise Value/Ebtida.

[2] Elemento de avaliação que mede a geração de caixa nas operações da empresa, antes que seja afetada pelos encargos financeiros e débitos contábeis. (BOVESPA, 2007)

### 1.5.2 Análise Técnica

Dentre os dois tipos de análises existentes, a análise técnica é a mais utilizada pelos investidores pessoa física, atraídos por sua aparente simplicidade, pois é basicamente a análise do gráfico histórico de preços de uma ação. Conforme os princípios desta análise, nestes gráficos já são descontados todas as outras informações relevantes que influenciam no preço da ação, dispensando a necessidade de avaliação de outras informações (MATSURA, 2007), permitindo ao investidor buscar a ação que tem maior potencial de ganhos. O objetivo principal dos praticantes desta análise é identificar a tendência de curto prazo do mercado e, por meio de gráficos e padrões, conceberem regras que possibilite a aplicação desta tendência (PAULOS, 2004). A análise técnica, por estudar tendências, baseia-se em algumas regras gerais, como, por exemplo, a Teoria de Elliot (MATSURA, 2007).

Ralph Nelson Elliot mostrou, em 1939, uma teoria em que os preços das ações movimentavam-se em ondas, onde na situação mais comum o mercado subia em cinco ondas distintas e caia em três ondas diferenciadas, por motivos obscuros, de natureza sistêmica ou psicológica. (PAULOS, 2004). Além disso, Elliot também descreveu que devido aos níveis do mercado, cada ciclo ou onda fazia parte de outra onda maior, que contém diversas ondas ou ciclos menores. Seguindo esta teoria, o resultado é um questionamento ao qual os investidores têm que encontrar o melhor momento de compra ou venda, identificando em qual ponto da onda se encontram no momento.

Elliot também demonstrou que a amplitude destas ondas poderia ser mensurada em função dos números da série de Fibonacci (1, 1, 2, 3, 5,8...,) onde cada número sucessivo na seqüência representa o resultado da soma dos dois anteriores. Elliot escreveu que uma vez determinados os níveis de suporte e resistência é possível definir os níveis de oscilações da tendência, através da razão da série de Fibonacci.

Hoje, no estudo das movimentações do mercado de ações, a relação da seqüência de Fibonacci utiliza a constante resultante da divisão dos números da série por seus respectivos antecessores que resulta em uma razão próxima a 62%, sendo o inverso deste número 1, 618. Ainda nesta teoria, a divisão de um número pelo seu segundo antecessor resulta em uma razão próxima a 38%. A partir destas constantes, foram identificadas porcentagens que,

teoricamente, seguem esta seqüência. Essas porcentagens são: 0% - 38,20% - 50% - 61,80% - 100%, traçadas com base em pontos de máxima ou mínima. Por exemplo, a análise do ativo PETR4 no movimento conforme o gráfico 1.2, mostra a correção do movimento de baixa do ativo, onde se respeitando a primeira resistência da série de Fibonacci (38,20%) houve o rompimento da resistência, criando um novo suporte para a tendência (a). Após esta oscilação o novo ponto de resistência seria a faixa de 50% ( R$ 67,22), sendo em seguida testado pelo mercado (b), não confirmando o rompimento da resistência. No movimento seguinte (c), houve o rompimento da resistência da mesma forma da oscilação anterior, criando um novo ponto de suporte e uma nova resistência. Esta nova resistência conforme as teorias de Elliot, estaria localizada na faixa de 61,80% (R$ 72,78). É possível observar que este valor de resistência foi "testado" pelo mercado, mas não confirmou seu suporte no primeiro movimento (d). Já no movimento posterior (e), esta resistência foi rompida de forma a determinar um novo ponto de suporte, na faixa dos R$ 72,78, confirmado no movimento de queda parcial (f). Após este movimento, seguindo as teorias de Elliot, o último ponto de resistência desta onda é no percentual de 100%, onde o preço da ação pode chegar a R$ 90,77.


**GRÁFICO 1.2** – ANÀLISE DE ATIVO COM USO DE SÉRIE DE FIBONACCI
**Fonte: RIOS, 2008.**

Já a Teoria de Charles Dow Jones, um dos primeiros grafistas dos Estados Unidos, identifica quatro fases que se repetem na formação dos preços. Para Dow Jones a opinião das massas comanda as oscilações do preço no mercado através das pressões de compra e venda e, além disso, as médias descontariam tudo (ELDER, 2005). Seguindo esta teoria, foi criado um índice, chamado de Dow Jones *Average* utilizado na Bolsa de Nova York, correspondendo no Brasil ao IBOV (Índice Bovespa). Segundo Deschatre (1997), este índice representa a média do índice das ações de maior representatividade no mercado.

De acordo com esta teoria, citado por Deschatre (1997), as médias descontam tudo, pois as mudanças diárias de preços refletem as emoções e julgamentos dos participantes do mercado. O autor defende que há uma relação entre os preços e os volumes operados, sendo que o volume cresce com a elevação do preço e decresce com a diminuição dele e que o preço das ações são o que determinam a tendência de oscilação. Além disso, o mercado possui 3 movimentos, sendo os movimentos primários, secundários e menores, respectivamente de longo, médio e curto prazos. Neste estudo, também são analisadas as médias setoriais que têm que confirmar a média global: Por exemplo, para confirmar a tendência de queda, enquanto a média global IBOV estiver caindo a média de outros setores devem estar no mesmo processo.

Como citado por Cavalcante; Misumi; Rudge (2005), as fases da teoria de Dow Jones são: A fase de **Acumulação**, onde os investidores muito bem informados adquirem títulos. ***MARK UP*** é onde compradores informados adquirem ações, elevando o preço da mesma. **Distribuição** é a fase onde a massa de investidores se interessa pelo ativo e compra ações, vendidas pelos investidores que adquiriram estas na fase de acumulação. Após o período de distribuição, ocorre a fase de **Liquidação**, em que a massa de investidores vende os títulos, no julgamento de terem comprado as ações por um valor elevado, não compensando o investimento. Em seguida ocorre novamente a nova fase de acumulação, fechando o ciclo.

Deschatre (1997) mostra baseado na teoria de Dow, que o mercado de capitais possui duas grandes fases: a de alta, onde os preços ganham forças, com sucessivas subidas e a fase de baixa, com sucessivas descidas. Ambas com reações correspondentes nos períodos mais curtos.

#### 1.5.2.1 Tipos de gráficos e indicadores

Cada tipo de gráfico possui características de visualização e fornecem informações diferenciadas ao analista. Conforme Cavalcante, Misumi e Rudge (2005), os principais gráficos utilizados na análise técnica são: Gráfico de Linhas e Barras; Indicador Relativo de Força; Média Móvel; Gráfico de volumes; Gráfico de Posições em Aberto; Gráfico Ponto e Vírgula; Gráfico de Força Relativa; Gráfico de Velas; Indicador de Avanço e Declínio; Gráfico de Preço-Quantidade; Análise de Torque.

Estes instrumentos visam permitir ao investidor avaliar tendências em séries de preços ou ações, a natureza dos investidores do mercado, a natureza cíclica das oscilações de preços e a importância dos volumes negociados. Os gráficos permitem a identificação de linhas de preços mínimas (suporte) e máximas (resistência). Os preços testam estas áreas e quando as rompem, quase sempre ocorrem as mudanças de tendências (CAVALCANTE; MISUMI; RUDGE, 2005). Para Elder (2005) **Suporte** é o nível de preço em que as compras têm força suficiente para interromper ou reverter a tendência de baixa. **Resistência** é o nível de preço em que as vendas têm força suficiente para interromper ou reverter a tendência de alta. Segundo Elder (2005) a força destas duas zonas dependem de sua intensidade, altura e do volume de negociações nelas realizadas.

### 1.6 Softwares Disponíveis

Como a análise técnica é baseada em indicadores estatísticos, a maior parte dos softwares encontrados no mercado já apresenta a maioria dos indicadores conhecidos (DESCHATRE, 2005). Atualmente, estão disponíveis no mercado global diversas ferramentas e softwares que visam auxiliar os investidores fundamentalistas e técnicos na tomada de decisão, sendo que para Deschatre (2005), para selecionar o melhor software os pontos a serem observados são a capacidade de armazenamento de séries históricas de acordo com o número de ações; as ferramentas disponíveis para análise gráfica; o acesso automático a base de dados; a possibilidade de controle da carteira com relatórios de avaliação de performance; as cotações com valores máximos, mínimos, fechamentos e volumes

negociados no dia e a capacidade de atualização automática dos direitos de acionistas e dos respectivos reflexos nas séries históricas.

### 1.7 Mercado Financeiro X *Data Mining*

Como citado por Bruni e Fama (1998), seguindo conceitos da Teoria de Finanças e analisando a eficiência dos mercados, os preços de ativos, ou seja, a operação de compra e venda de ativos resultaria em um valor presente nulo, onde não é possível prever comportamento do mercado. Porém, a previsibilidade do comportamento dos preços no mercado de ativos vem ao longo do tempo sendo estudada por acadêmicos com sofisticadas técnicas na tentativa de predição de preços futuros, contrariando esta teoria, ao obterem resultados significativos em suas pesquisas. Dentre as técnicas estudadas encontram-se inclusive, as técnicas de mineração de dados.

Conforme citado por Deschatre (1997), as fases e ações do mercado podem ser visualizadas através de gráficos e com auxílio do computador. A análise é feita com maior velocidade e flexibilidade, permitindo a coleta e manipulação de dados de forma rápida. Entretanto, os indicadores utilizados na análise técnica não raras vezes são contraditórios entre si, cada um deles funcionando melhor em mercados de tendências e outros em mercados estáveis (horizontais) (ELDER, 2005).

Segundo Matsura (2007), a crescente complexidade dos instrumentos de negociação do mercado financeiro, assim como o acesso a novas tecnologias de processamento da informação, estimula o desenvolvimento de novos sistemas de análise e operação; inclusive com o uso de técnicas de Inteligência Artificial (ZANETI; ALMEIDA, 1998).

Através da aplicação de técnicas de *Data Mining* neste contexto, mais especificamente no mercado à vista, pretende-se a identificação de padrões de comportamentos hoje despercebidos pelos investidores, e, com isso, determinar a tendência futura dos ativos do mercado à vista.

O capítulo a seguir demonstra como é estruturado o processo de descoberta de conhecimento e as técnicas que serão utilizadas na exploração do contexto abordado.

# 2 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS(KDD)

O mercado de capitais, foco deste estudo, gera diariamente uma enorme quantidade de dados nas análises e transações dos diversos mercados. Desta forma, caracteriza-se como uma área de aplicação para tecnologias da informação, como a descoberta de conhecimento.

Conforme Rezende (2005), o armazenamento de dados em sistemas computacionais tornou-se uma tarefa essencial para instituições de uma forma geral. Fator este devido principalmente à queda dos custos de armazenamento, a rápida automatização das empresas, versatilidade no acesso a informação e a Internet. Com isso, durante os últimos anos tem se verificado um crescimento substancial na quantidade de dados armazenados, espalhados em bases de dados, *datawarehouses*, e outros tipos de repositórios de dados nos diversos segmentos da sociedade, o que torna os métodos de análises manuais de dados incompatíveis com a realidade vivida (AURÉLIO, VELLASCO; LOPES, 1999). Entretanto, conforme diversos autores da área, grandes quantidades de dados representam um potencial cada vez maior de informação e, portanto, um diferencial competitivo e o maior bem das instituições em nossa sociedade.

Para Goldschmidt e Passos (2005), examinar e interpretar de forma correta grandes volumes de dados brutos exige talento e capacidade técnica não trivial até mesmo para especialistas das áreas envolvidas, pois como citado por Rezende (2005), o conhecimento raramente é obtido de forma direta. Logo, diante deste cenário, surge a necessidade de explorar estes dados para extrair informação - conhecimento implícito, e utilizá-la no âmbito do problema (AURÉLIO, VELLASCO; LOPES, 1999).

Segundo Rezende (2005), as instituições têm buscado na tecnologia recursos que agreguem valor aos seus negócios, seja agilizando operações, suportando ambientes ou viabilizando inovações nos diversos segmentos da sociedade. Para Carvalho (2000), é inevitável o aparecimento de técnicas e ferramentas de recuperação muito mais eficazes do que as técnicas existentes, a fim de atender às necessidades dos usuários.

Este é o foco da Descoberta de Conhecimento em Base de Dados ("*Knowledge Discovery in Databases*" – KDD) (CARVALHO, 2000), considerada por Goldschmidt e Passos (2005) o tópico de maior relevância para a área da informática atualmente.

## 2.1 Processo de Descoberta de Conhecimento (KDD)

O processo de KDD é um conjunto de diversas etapas operacionais (GOLDSCHMIDT, 2005), conforme representado na figura 2.1.

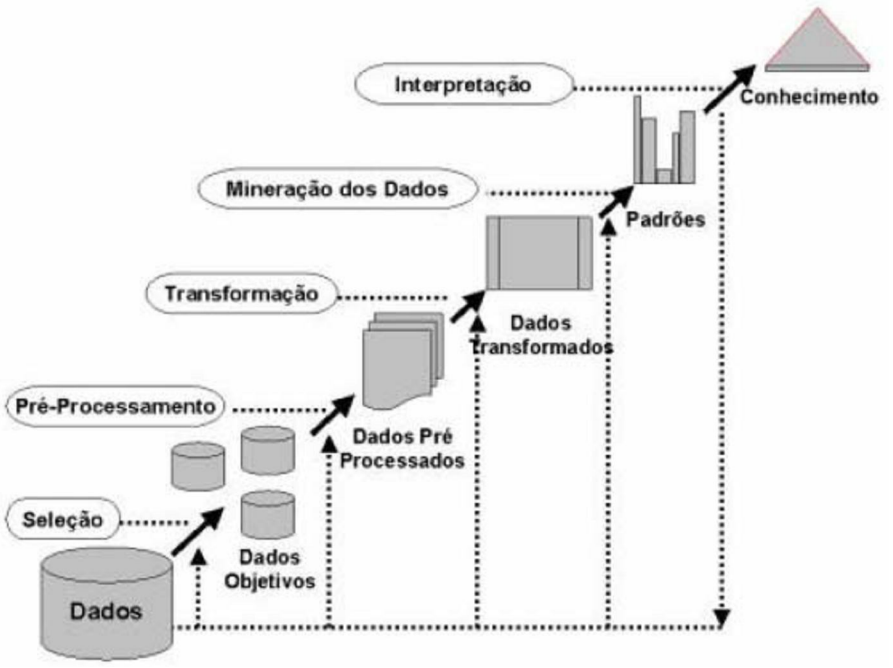


**FIGURA 2.1**: FASES DO KDD.
**Fonte:** AURÉLIO, VELLASCO; LOPES, 1999, p.21.

Primeiramente, antes da execução das etapas do processo mostrado acima, é necessário analisar e definir os objetivos a serem atingidos com a extração do conhecimento no contexto da aplicação (GONCHOROSKI, 2007). Nesta fase, são definidos itens como a aderência entre o escopo de aplicação e a tecnologia de KDD e a relação custo-benefício da aplicação desta tecnologia. Analisadas essas questões e julgando-se KDD a solução para o problema, a etapa de preparação de dados pode ser iniciada (CARVALHO, 2000)

Assim, os dados passam pela etapa de pré-processamento, cujo objetivo é definir os dados úteis à extração de conhecimento, validá-los e assim, reduzir a complexidade do problema. Existem nesta fase 3 sub-etapas, conhecidas como limpeza, seleção e transformação de dados, onde são identificadas inconsistências e os atributos mais significativos são selecionados (CARVALHO, 2000; CARVALHO, 2005).

Seleção é a fase onde os dados realmente úteis para o objetivo do processo são identificados. Esta fase pode ter, conforme Goldschmidt e Passos (2005) dois enfoques: a escolha de atributos ou a escolha de registros a serem utilizados, que englobam operações como simplificação, eliminação e agregação de dados.

A fase de limpeza tem como objetivo retirar do conjunto de dados selecionados as amostras que estejam com dados incompletos, ruidosos ou inconsistentes. São operações comuns nesta fase a verificação da consistência das informações, preenchimento ou eliminação de valores desconhecidos ou redundantes, a correção de possíveis erros e a retirada de valores fora do escopo do processo. (REZENDE, 2005).

A fase de transformação dos dados busca tornar os dados pré-processados na etapa anterior compatíveis com as entradas dos algoritmos de mineração de dados. Nesta etapa, por exemplo, as variáveis contínuas podem ser codificadas em variáveis categóricas, pode ocorrer a criação de novas variáveis, conversões de tipos, etc (REZENDE,2005;CARVALHO,2000).

Cabe salientar que conforme citado por Rezende (2005), as transformações descritas devem ser realizadas de forma a garantir que as informações presentes nos dados brutos continuem presentes nas amostras, para que os modelos finais sejam compatíveis com a realidade da aplicação.

A próxima etapa, a mineração de dados, é composta pela garimpagem dos dados. Nesta etapa são definidos o algoritmo (técnica) e a tarefa a serem utilizadas para a exploração eficiente da massa de dados. A escolha da tarefa de mineração de dados varia de acordo com os objetivos da aplicação. Estas tarefas podem ser agrupadas em atividades preditivas e descritivas.

As atividades preditivas consistem na generalização de dados históricos com respostas conhecidas para o reconhecimento de classes de um novo exemplo. As principais tarefas preditivas são as classificações e as regressões. A classificação consiste na predição de um valor categórico, enquanto a regressão trata de um valor contínuo, buscando prever o valor da próxima ocorrência.

As atividades descritivas buscam a identificação de valores intrínsecos do conjunto de dados, sem classe especificada previamente. Algumas das tarefas descritivas são as regras de associação, *clustering* e a sumarização (REZENDE, 2005).

A escolha do algoritmo está intimamente ligada à complexidade da solução e seus objetivos propostos (REZENDE, 2005). Para Goldschmidt e Passos (2005), existem diversos tipos de técnicas e algoritmos a serem utilizados para mineração de dados, sendo estes subdivididos em técnicas tradicionais, técnicas específicas e técnicas híbridas, onde sua escolha depende muitas vezes da tarefa a ser realizada.

As técnicas tradicionais são as tecnologias independentes do contexto da mineração de dados, geralmente produzindo bons resultados nas aplicações de descoberta de conhecimento. Como exemplos destas tecnologias podem ser citados redes neurais artificiais (RNA), lógica nebulosa, árvores de decisão, algoritmos genéticos e a estatística.

As técnicas específicas são aquelas desenvolvidas para atender especificamente às tarefas de descoberta de conhecimento. Como exemplo típico na área, existe o algoritmo *apriori*, desenvolvido especificamente para a descoberta de associação.

As técnicas híbridas, geralmente mais eficientes e com menos deficiências, podem inclusive ser consideradas sistemas por combinar mais de uma técnica para a solução do

problema. Estas técnicas são classificadas de acordo com a forma de integração entre as tecnologias utilizadas, sendo estas definidas como híbridos seqüenciais, auxiliar e incorporado.

Para Goldschmidt e Passos (2005), a etapa seguinte, interpretação, compreende a tarefa de avaliação dos resultados pelo especialista do domínio da aplicação e pelo especialista em descoberta de conhecimento. Conforme Rezende (2005), o conhecimento obtido a partir da aplicação das técnicas de mineração de dados pode ser utilizado como apoio a tomada de decisão de um ser humano ou por meio de um sistema especialista. Além disso, este conhecimento deve ser avaliado quanto sua representatividade e validade como um novo conhecimento através de medidas de desempenho e qualidade.

Após a análise dos resultados obtidos, caso estes não atinjam o objetivo proposto, o processo de extração de mineração pode ser ajustado e repetido. (REZENDE, 2005). Como a parte principal do processo de KDD envolve diversas técnicas e tarefas complexas, estes processos serão estudados de forma mais detalhada a seguir.

## 2.2 Mineração de Dados

Mineração de Dados é um campo de pesquisa interdisciplinar que mescla conceitos de estatística, inteligência artificial, computação paralela e banco de dados. Seu principal desafio é a extração de conhecimento implícito em grandes e densas bases de dados possuindo vasta aplicabilidade em várias áreas, incluindo suporte a decisões, formulação de estratégias de marketing e previsões na área financeira (CARVALHO, 2005).

Conforme Carvalho (2005), Mineração de Dados ou *Data Mining* (DM) é um conjunto de técnicas reunidas da Estatística e da Inteligência Artificial, com objetivo de descobrir conhecimento novo que esteja “escondido” em grandes massas de dados. Já Harrison (1998) (Apud AURÉLIO; VELLASCO; LOPES, 1999) define DM como um processo de exploração e análise de uma grande massa de dados que através de meios automáticos ou semi-automáticos, objetivando a descoberta de padrões e regras significativas.

O volume, complexidade e a peculiaridade dos eventos tratados nesta etapa impõem diversas limitações ao uso somente das técnicas estatísticas. Algumas limitações são a inter-relação entre os atributos dos domínios de aplicação que podem comprometer o resultado da análise, o alto grau de conhecimento estatístico exigido para aplicação de técnicas estatísticas, os métodos estatísticos não manipulam de forma aceitável valores simbólicos, incompletos ou inconclusivos e por fim, os métodos estatísticos tornam-se computacionalmente caros quando aplicados a grandes bases de dados. Devido a estas dificuldades, a mineração de dados utiliza técnicas de diversas áreas de conhecimento como inteligência artificial, estatística e banco de dados de forma integrada, tratando estas restrições (SILVA, 2006).

A extração de conhecimento de grande base de dados e a utilização de *data mining* para a realização desta tarefa está presente em diversos ramos do mercado, (CARVALHO, 2000) e em diversos domínios de aplicações como *marketing*, análises corporativas, astronomia, medicina, biologia, entre outros (AURÉLIO; VELLASCO; LOPES, 1999).

Desta forma, são identificadas diversas tarefas de KDD, dependentes do domínio de aplicação, destinadas cada uma a extração de um tipo diferente de conhecimento. Conforme Carvalho (2005), basicamente cinco tarefas ou classes de aplicação abrangem didaticamente todas as áreas de mineração, permitindo assim uma visão geral da função de cada grupo, sendo elas a classificação, estimativa, previsão, análise de afinidade e análise de agrupamentos (Clusterização). A seguir será abordada a tarefa utilizada neste estudo.

### 2.2.1 Classificação

A Classificação é a tarefa de mineração de dados mais comum e consiste em identificar e classificar padrões ou grupos. Nesta tarefa, o dado analisado é comparado com outros que pertençam a classes ou grupos já definidos através de uma métrica de diferença entre eles. É uma tarefa comumente utilizada na identificação de transações financeiras legais, ilegais ou suspeitas, por exemplo (GIUDICI, 2003). Depois de identificada a relação entre os elementos, esta função pode ser aplicada a novos registros, prevendo a relação destes com as respectivas classes (GOLDSCHMIDT; PASSOS, 2005).

São exemplos da tarefa de classificação: descoberta de padrões de consumidores, classificar pedidos de créditos como de baixo, médio e alto risco; identificar fraudes em seguros, etc. Como exemplo, analisando a tabela 2.1 pode-se identificar regras de relacionamentos entre os possíveis compradores de um produto, baseando-se em dados como sexo, país, idade e histórico de transações efetuadas ou não.

**Tabela 2.1:** Entrada de dados para tarefa de classificação.

| SEXO | PAÍS | IDADE | COMPRAR |
|---|---|---|---|
| Masculino | Brasil | 25 | Sim |
| Masculino | Argentina | 21 | Sim |
| Feminino | Brasil | 23 | Sim |
| Feminino | Argentina | 34 | Sim |
| Feminino | Brasil | 30 | Não |
| Masculino | Uruguai | 21 | Não |
| Masculino | Uruguai | 20 | Não |
| Feminino | Uruguai | 18 | Não |
| Feminino | Brasil | 34 | Não |
| Masculino | Brasil | 55 | Não |

**Fonte:** Adaptado de Goldschmidt e Passos, 2005, p.70.

O conhecimento descoberto é frequentemente representado na forma de regras **SE-ENTÃO**. Seguindo esta idéia, as regras geradas são as seguintes:

**Tabela 2.2:** Regras geradas após a tarefa de classificação.

| Regras Geradas |
|---|
| Se (PAÍS = Uruguai) então COMPRAR = Não. |
| Se (PAÍS = Argentina) então COMPRAR = Sim |
| Se (PAÍS = Brasil e IDADE ≤ 25) então COMPRAR = Sim. |
| Se (PAÍS = Brasil e IDADE >25) então COMPRAR = Não. |

**Fonte:** Adaptado de Goldschmidt e Passos, 2005, p.71.

Conforme as regras geradas, se o país do comprador for Uruguai o histórico não mostra efetivação nas transações. Caso o país do comprador seja Argentina, o histórico da base de dados analisada mostra a efetivação das transações. No caso do país do comprador for o Brasil, ainda existe a necessidade da avaliação do fator IDADE para a identificação de padrões

de relacionamento. Caso a idade seja acima de 25 anos, o histórico analisado não mostra a efetividade das transações. Caso a IDADE seja igual ou inferior a esta idade, o histórico mostra a efetividade das transações. Com estes dados, campanhas de marketing, logística e planejamentos estratégicos podem ser definidas.

Conforme o teorema de NFL (*No Free Lunch Theorem*), todos os algoritmos de classificação possuem a mesma importância em qualquer problema de classificação. Logo, a cada nova aplicação todos devem ser testados, identificando os de melhor desempenho (GOLDSCHMIDT; PASSOS, 2005).

Uma vez selecionada uma hipótese (classificador) esta pode ser muito específica para o conjunto utilizado. Entretanto, existem casos onde o classificador fica "viciado" no ambiente de treinamento, efeito causado por um fenômeno chamado de *overfitting*. Por outro lado, quando não há "aderência" entre o classificador e o conjunto de treinamento, ocorre um *underfitting*, geralmente justificado por parametrizações inadequadas do algoritmo, como por exemplo, um número insuficiente de neurônios em uma rede neural (GOLDSCHMIDT; PASSOS, 2005; REZENDE, 2005).

Algumas das ferramentas utilizadas nesta tarefa são as redes neurais artificiais, estatísticas e algoritmos genéticos, C4.5, K-NN, *Back-propagation* e classificadores Bayesianos (GOLDSCHMITD;PASSOS, 2005).

O uso de técnicas de mineração de dados e inteligência artificial no mercado financeiro para determinação de estratégias e prever as tendências de alta e baixa já é realidade desde 1986 nos Estados Unidos. Primeiro, através da aplicação de sistemas baseados em regras heurísticas, com o uso de sistemas especialistas. Após, com o uso de análise estatística e, finalmente, com o uso de redes neurais artificiais. Todas as opções com suas respectivas vantagens e desvantagens (CARVALHO, 2005).

Como citado por Montgomery e Ludwig (2007), as redes neurais com aprendizado não supervisionado, como a rede de Kohonen tem por finalidade a classificação de dados pelo reconhecimento de padrões. Porém, como citado por (ABELÉM, 1994), as redes de aprendizado supervisionado, especialmente as redes que utilizam o algoritmo de retro-propagação tem mostrado resultados significativos na predição de valores no mercado

acionário. Além disso, conforme o mesmo autor, as redes neurais artificiais vêm mostrando superioridade com relação aos métodos estatísticos convencionais no tratamento e identificação de padrões com dados não lineares e muito ruidosos.

Alinhado com os objetivos do estudo e como a tecnologia de redes neurais está se mostrando cada vez mais forte, destacando-se no cenário econômico, e com base nos estudos analisados (ZANETI; ALMEIDA, 1998) (LUDWIG; MONTGOMERY, 2007), foi definida como uma das tecnologias a ser utilizada. Além disso, esta tecnologia contempla as tarefas pretendidas para a identificação de padrões no mercado financeiro, como classificação e previsão.

Entretanto, devido ao alto grau de abstração nas justificativas de suas respostas, faz-se necessária a comparação dos resultados obtidos com alguma outra tecnologia cujo funcionamento seja claro. Visando atender a este requisito, será também abordado neste estudo o uso da técnica de árvores de decisão para classificação de ativos e comparação dos resultados obtidos, devido a sua fácil visualização do processo de classificação. Ambas as técnicas de aprendizado serão vistas no próximo capítulo.

# 3 CLASSIFICADORES

Dentro da Mineração de Dados a classificação de padrões é a área mais estudada. Os padrões de classificação possuem atributos de dois tipos: os preditivos e os objetivos. Os atributos objetivos tratam de variáveis categóricas, representando classes previamente definidas, conhecidas na fase de treinamento. Já os atributos preditivos buscam identificar a que classe um novo padrão pertence. Na fase de treinamento, o objetivo deste tipo de padrão de classificação é identificar e aprender as relações entre os atributos preditivos e objetivos. Visando a comparação da classificação entre dois tipos de classificadores, a seguir são apresentadas as duas técnicas utilizadas na tarefa de classificação neste estudo, a técnica de redes neurais artificiais e árvores de decisão.

## 3.1 Redes Neurais Artificiais

As redes neurais artificiais (RNAs) são técnicas de Inteligência Artificial, hoje empregadas em diversas áreas de conhecimento e estudo, como neurociência, matemática, física, economia, engenharias e ciência da computação, focadas em tarefas como a previsão de valores, padrões, classificações de clientes, e outras. Estas estruturas têm como vantagem a capacidade de aprender com os dados de entrada, e por isso, tem o interesse de pesquisadores do mundo todo em seu funcionamento e evolução (LUDWIG; MONTGOMERY, 2007).

Diversos trabalhos nesta área foram desenvolvidos, alguns inclusive com resultados expressivos, como por exemplo, o estudo desenvolvido por BERGENSON e WUNSCH (Apud ZANETI; ALMEIDA, 1998), onde com o uso de um sistema composto por redes neurais híbridas, obtiveram um retorno de 660% em 25 meses com uma aplicação de US$ 10.000 entre janeiro de 1989 e janeiro de 1991. Um dos aspectos negativos deste método, segundo os autores, foi o grande número de horas gasto para a construção do modelo. Estes fatos ressaltam a existência de um vasto campo de estudo acerca da comparação do desempenho de modelos preditivos para previsão de ativos em bolsas de valores.

### 3.1.1 Vantagens e Desvantagens

A rede neural é uma estrutura com processos em paralelos, que possui habilidade de generalização, ou seja, produz saídas adequadas para as entradas que não faziam parte do seu treinamento. Porém na prática, as redes não conseguem fornecer soluções para problemas muito complexos trabalhando sozinhas (LUDWIG; MONTGOMERY, 2007).

Uma vantagem das redes neurais é a possibilidade de se projetar a rede para modificar seus pesos sinápticos em tempo real. Além disso, a estrutura da rede neural dá a ela o potencial de ser tolerante a falhas, devido ao armazenamento distribuído das informações na rede. A principal desvantagem está no seu processamento que é considerado uma “caixa preta” pela dificuldade em determinar “como” a rede chegou ao resultado, seja em uma rede de uma ou mais camadas. A forma comumente utilizada para verificar o correto funcionamento da rede é pela análise do erro médio quadrático (LUDWIG; MONTGOMERY, 2007).

### 3.1.2 Conceitos Básicos

Uma *RNA* é uma ferramenta computacional que busca simular o comportamento do cérebro humano (FERNANDES, 2005). Para o desenvolvimento de um modelo computacional de comportamento inteligente, neurologistas e pesquisadores da inteligência

artificial propuseram uma rede altamente interconectada de nódulos (ou neurônios) (AURÉLIO; VELLASCO; LOPES, 1999). Esta estrutura tem a capacidade de aprendizado, generalização, associação e abstração. As RNA tentam aprender padrões diretamente dos dados através de um processo de repetidas apresentações dos dados à rede, ou seja, por experiência. Dessa forma, uma RNA procura por relacionamentos, constrói modelos automaticamente, e os corrige de modo a diminuir seu próprio erro. Os componentes principais de uma RNA conforme Rumelhart (1996) (apud PORTUGAL; FERNANDES, 1996), são os neurônios ou unidades de processamento, a função de saída; as regras de ativação e propagação, a estrutura de ligação entre os neurônios (rede) e o ambiente ao qual esta metodologia será empregada (dados utilizados).

#### 3.1.2.1 O Neurônio Artificial

O cérebro humano é composto por mais ou menos $10^{11}$ neurônios, e apesar de não existirem dois neurônios iguais, eles possuem características comuns (ABELÉM, 1994). Como citado por Carvalho, (2005), a estrutura básica do funcionamento cerebral é a célula nervosa, também chamada de Neurônio, que tem como função gerar sinais elétricos e transmiti-los a outras células. O neurônio humano tem sua estrutura composta por dendritos, corpo, axônio e colaterais.

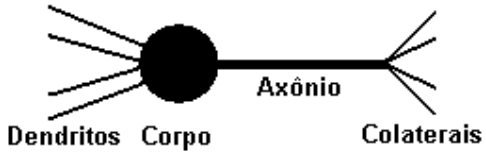


**Figura 3.1**: O Neurônio humano.
**Fonte:** Adaptado de Carvalho, 2005, p. 94.

O núcleo da célula, ou soma, está localizado no corpo da mesma. Conectados ao corpo da célula existem fibras nervosas chamadas *dendritos*. Estendendo-se do corpo da célula existe uma única fibra nervosa mais grossa chamada *axônio*, da qual surgem ramificações e sub-ramificações chamadas de *colaterais*, onde se encontram as junções sinápticas ou *sinapses*. Sinapse é a conexão entre um colateral e um dendrito. Os impulsos elétricos são enviados do corpo para os colaterais através do axônio. Porém a ligação entre os

neurônios não é direta. O impulso elétrico é transformado em uma codificação química por substâncias liberadas pelo neurônio transmissor através de neurotransmissores e receptores presentes nas sinapses. O efeito é um aumento ou uma queda no potencial elétrico no corpo da célula receptora. Se este potencial alcançar o limite de ativação da célula, um pulso é enviado através do *axônio*. Diz-se então que o neurônio está ativo.

O neurônio artificial foi elaborado para imitar as características de um neurônio biológico. Basicamente, um conjunto de entradas são aplicadas ao neurônio artificial, cada uma representando a saída de outros neurônios, ou seja,os dendritos. Cada entrada tem um valor (x) que é multiplicada por um peso correspondente (Wij), gerando entradas ponderadas, simulando as sinapses de um neurônio natural. Em seguida todas estas entradas ponderadas são somadas, obtendo-se um valor NET que será comparado com o valor limite para ativação do neurônio (F). Caso este valor alcance o valor limite de ativação do neurônio, ele se ativará, caso contrário ele ficará inativo.

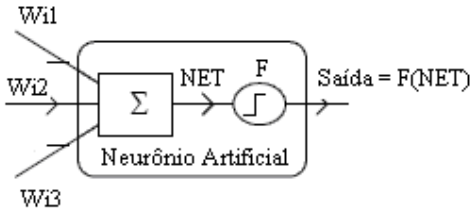


**Figura 3.2:** Estrutura do neurônio artificial.
**Fonte:** AURÉLIO, VELLASCO e LOPES, 1999.

#### 3.1.2.2 Estado de Ativação

O estado de ativação do sistema especifica o que está sendo representado na rede em um instante *t* qualquer. Estes valores podem ser discretos, como por exemplo, valores de {0,1} ou {-1,0,1} ou valores contínuos no intervalo [0,1] ou [-1,1], que serão calculados pela regra de ativação.

#### 3.1.2.3 Função de Saída:

A função de saída é o mapeamento do estado de ativação em um sinal de saída, sendo o limiar que permite ou não a ativação do neurônio.

#### 3.1.2.4 Padrão de Interconexão

Pode ser representado por uma matriz de pesos W, onde um elemento Wij corresponde à interferência do neurônio Ui sobre o neurônio Uj. Nesta representação, pesos positivos representam a excitação ou o reforço na ativação do neurônio Uj. Da mesma forma, os valores de pesos negativos representam a inibição na função de ativação do neurônio Uj (FERNANDES, 2005).

Topologicamente, as redes neurais podem ser organizadas em camadas. Em termos gerais, as redes possuem uma camada de entrada que não possui neurônios, tendo apenas um número de nós igual ao número de sinais de entrada da rede e não realiza funções computacionais. Após, uma, nenhuma ou diversas camadas ocultas, com um ou mais neurônios ocultos. Estes neurônios capacitam a rede a extrair estatísticas e realizam as funções computacionais da rede. Entre as camadas da rede podem existir conexões entre camadas (inter-camadas) e conexões entre neurônios de mesma camada (intra-camadas). E, por fim, uma camada de saída que contém um número de neurônios igual ao número de sinais de saída da rede.

Com o ajuste dos pesos sinápticos, a rede neural pode memorizar as relações entre as entradas e saídas e, assim, assumindo uma característica de memória associativa (LUDWIG; MONTGOMERY, 2007). Como Exemplo, toma-se o citado por Portugal e Fernandes (1996).

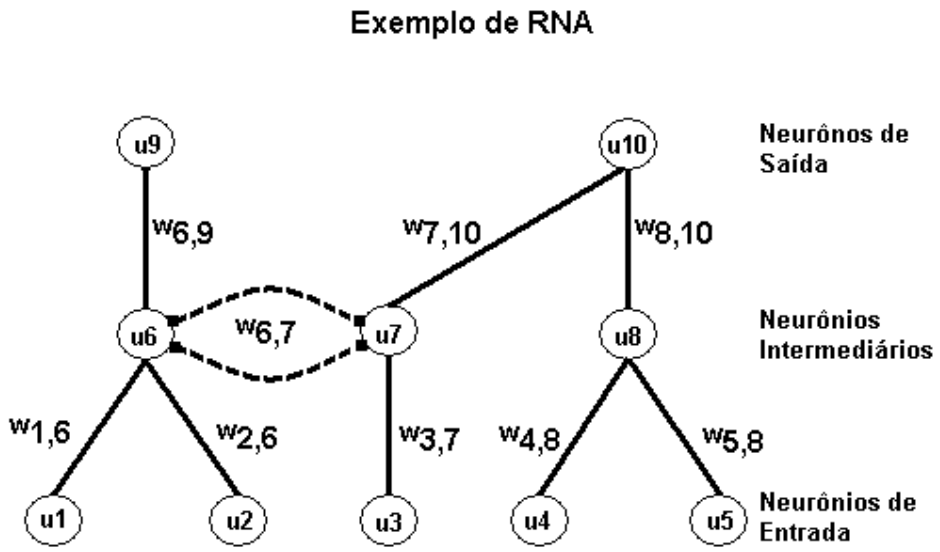


**Figura 3.3:** Exemplo de Rede Neural Artificial.
**Fonte:** PORTUGAL; FERNANDES, 1996.

Na figura acima, os neurônios u1, u2, u3, u4, u5 são os neurônios de entrada da rede. Os neurônios u9, u10 são os neurônios de saída. Os neurônios u6, u7, u8 são os neurônios da camada intermediária (camada oculta). Existem conexões inter-camadas, ou seja, conexões entre neurônios de camadas diferentes, como por exemplo, a ligação entre os neurônios u7 e u10 , u1 e u6, u4 e u8. Além deste tipo de conexão, no exemplo é mostrada um tipo de ligação entre neurônios de mesma camada (entre os neurônios u6 e u7), sendo chamada de conexão intra-camada. A ligação entre os neurônios u6 e u7 formam um ciclo entre os neurônios, sendo conhecida como uma ligação recorrente. Este tipo de conexão forma um ciclo voltando o sinal para o neurônio que foi ativado primeiro. Esta ligação poderia ocorrer com neurônios entre camadas diferentes. A presença ou não deste tipo de ciclo na rede é um fator de classificação das redes em redes cíclicas e acíclicas.

Quando, por exemplo, o neurônio u10 é ativado, o valor desta ativação é determinado pelos parâmetros das ativações de u3, u4, u5, u7, u8 e os pesos w4,8, w5,8, w3,7, w7,10, w8,10.

#### 3.1.2.5 Regra de Propagação

É a regra pela qual cada neurônio calcula sua nova ativação. Basicamente é a soma de todas as entradas dos pesos (net), dos neurônios Uj, após a multiplicação de cada valor de entrada pelos pesos das respectivas conexões (Wij). A regra de propagação fica completa com a utilização do limiar citado na função de saída.

#### 3.1.2.6 Regra de Ativação

É a regra que calcula o valor do neurônio no instante t, utilizando as entradas líquidas (net). Geralmente esta função tem o formato $a_i(t+1) = f(a_i(t) , net_i(t))$, onde f é a função de ativação ou função limiar. As quatro funções de ativação mais utilizadas são linear, rampa, degrau e sigmóide. Além destas, a função de ativação pode ser outra função que simule mais precisamente as características não lineares de neurônios biológicos.

#### 3.1.2.7 Regra de Aprendizado

Sendo considerada uma variante da regra de Hebb, que estabeleceu o princípio do aprendizado pela variação dos pesos das conexões, as regras de aprendizado são a forma em que a rede altera seu padrão de interconexão, alterando os pesos das conexões com base na experiência adquirida durante o aprendizado. Existem basicamente três formas de realizar esta tarefa: desenvolvendo novas conexões, ou com a perda das conexões existentes na rede ou ainda, com a modificação dos pesos das conexões existentes (PORTUGAL; FERNANDES, 1996).

Basicamente ocorrem somente as modificações dos pesos das conexões, pois em uma matriz de pesos ($w_{ij}$), as duas primeiras formas de modificação do padrão de interconexão podem ser simuladas. O desenvolvimento de novas conexões ocorre quando o peso de uma conexão de peso zero toma um valor positivo ou negativo. Da mesma forma, quando uma

conexão de peso positivo ou negativo recebe o valor zero , pode-se dizer que a conexão é desconectada (PORTUGAL; FERNANDES, 1996).

O conceito de aprendizado é definido dentro do campo da Inteligência Artificial como a habilidade de realizar tarefas novas que não podiam ser realizadas anteriormente (CARBONELL (1989), apud PORTUGAL; FERNANDES, 1996). As redes neurais podem ser classificadas quanto ao controle do aprendizado e conforme a arquitetura da rede.

As redes neurais, conforme o controle do aprendizado, podem ser classificadas em redes de aprendizado supervisionado e não-supervisionado. No aprendizado supervisionado, o treinamento da rede é dado com a apresentação da rede para pares de dados de entrada, onde a rede fornece os respectivos dados de saída. Após esta apresentação, ocorre o ajuste de pesos (sinapses) e níveis de bias até que a taxa de acerto seja considerada satisfatória. Este processo é repetido para todos os conjuntos de entradas e saídas do conjunto de treinamento da rede (LUDWIG; MONTGOMERY, 2007). No aprendizado não-supervisionado, a rede não recebe informações sobre as saídas desejadas, sendo treinada apenas com os valores de entradas, organizando os valores de entradas em grupos por classificação de características comuns dos elementos de entrada. A organização é feita por meio de processos de competição e cooperação entre os neurônios.

Considerando-se a arquitetura da rede, existem dois tipos principais de classificações, as amplamente conectadas e as redes com realimentação. As redes onde todas as saídas de uma camada de neurônios são conectadas a todos os neurônios da camada seguinte são consideradas redes amplamente conectadas (*fully connected*). Quando existem laços de realimentação na rede, ou seja, quando o sinal de saída se algum neurônio servir de entrada para um ou mais neurônios da mesma camada ou de camadas anteriores, a rede é considerada com realimentação (*feedback*). Estes laços, conforme Ludwig e Montgomery (2007) possuem grande impacto no aprendizado da mesma.

O treinamento de uma rede é processo interativo de ajuste de seus parâmetros livres, ou seja, dos pesos sinápticos e de bias com relação ao ambiente. Atualmente existem diversos algoritmos de treinamento (HAYKIN, 2001), onde se destacam aprendizagem competitiva (regra de Kohonen); a aprendizagem baseada em memória; a aprendizagem Hebbiana; a aprendizagem de Boltzmann e a aprendizagem por correção de erros (regra delta);

No aprendizado por correção de erros, são apresentados pares de treinamento à rede, correspondendo ao padrão de entrada e a resposta desejada. O erro (*E*) é obtido pela diferença entre a resposta desejada e a resposta obtida. A velocidade de correção, ou a taxa de aprendizado é uma constante positiva auxiliada por um método de gradiente descendente para corrigir os erros. O processo de correção e ajuste da rede ocorre até a satisfação da saída desejada, ou seja, o erro global de treinamento (média dos erros instantâneos de cada ciclo) atinja valor zero ou similar. Os pesos (w) são corrigidos com atualizações seguindo a relação:

Wnovo:= W anterior + ηE X

Onde: *X* é o valor de entrada do neurônio e o erro *E* é definido como: Resposta Desejada - Resposta Obtida (d - O); Além disso, a taxa de aprendizado $\eta$ é uma constante positiva, que corresponde à velocidade do aprendizado.

#### 3.1.2.8 Ambiente

Para o correto funcionamento das RNA, os possíveis padrões de entradas e saídas devem ser estabelecidos, sendo representados por uma função estocástica que varia dentro de um grupo de padrões de entrada no decorrer do tempo. As diversas RNAs possuem os mesmos conceitos e componentes, mas podem variar conforme a finalidade e a composição dos neurônios e as interligações entre eles.

### 3.1.3 Perceptron

A primeira Rede neural criada, a Perceptron de uma camada, é formada por uma camada única de neurônios de saída, diretamente conectados por pesos às entradas. A soma ponderada do produto entre pesos e entradas alimenta cada neurônio de saída, e se o resultado desta operação exceder um certo limiar (geralmente 0), o neurônio coloca o valor 1 na saída; se o resultado for inferior ao limiar, o neurônio coloca o valor -1 na saída (LUDWIG; MONTGOMERY, 2007). Esta rede pode ser treinada por um algoritmo de aprendizado

conhecido como regra-delta, que calcula o erro entre a saída obtida e a saída desejada. A partir do erro encontrado são ajustados novos pesos, até a obtenção da saída desejada ou similar. Os perceptrons de uma camada são capazes de aprender somente sobre problemas linearmente separáveis e se mostram incapazes de aprender a função XOR.

Por fim, a introdução de camadas internas entre as camadas de entrada e saída da rede perceptron de uma camada eliminaram o problema de aprendizado da função XOR.surgindo a Rede Perceptron Multi-camadas que consiste de múltiplas camadas de neurônios interconectadas, geralmente em forma *feedward*, ou seja, cada neurônio de uma camada tem conexões diretas aos neurônios da camada seguinte. Esta rede, por ser uma generalização da rede perceptron de uma camada tem o funcionamento descrito como uma seqüência de perceptrons

Em muitas citações os neurônios que compõe estas redes aplicam uma função chamada sigmóide como função de ativação (HAYKIN, 2001). Este tipo de rede é treinado de forma supervisionada, utilizando para aprendizagem o algoritmo conhecido como algoritmo de retropropagação de erro (LUDWIG; MONTGOMERY, 2007). Resumidamente, neste mecanismo os valores de saída são comparados com a resposta correta através de uma função de erro predefinida . Com esta informação, os pesos são ajustados por meio de um algoritmo, visando minimizar o valor da função de erro. Este processo se repete até a saída obtida ficar próxima do valor de saída desejado.

#### 3.1.3.1 O Algoritmo de Retropropagação de Erro

O aprendizado de retropropagação (*Back-propagation)* baseia-se na propagação retrógrada do erro para os níveis anteriores da rede, conforme o grau de participação que cada neurônio teve no nível posterior. Este algoritmo apresenta duas fases distintas. A primeira é o treinamento da rede e a segunda fase é a validação do resultado obtido. Nesta última etapa a rede deve reconhecer os padrões que foram treinados. Além disso, este tipo de algoritmo necessita da definição do sinal de erro e taxa de aprendizagem da rede. Visando garantir uma maior convergência no aprendizado da rede alguns fatores são atribuídos ao sinal de erro.

Resumidamente, este algoritmo pode ser descrito em cinco etapas (LUDWIG; MONTGOMERY, 2007).

Na primeira etapa são determinados valores aleatórios aos pesos sinápticos e níveis de bias uniformemente. Na etapa seguinte, são apresentados exemplos de treinamento em ciclos de um número de iterações definidas. Para cada exemplo, realiza-se a propagação dos sinais e a retropropagação dos erros.

Na terceira etapa, a propagação dos sinais inicia com a aplicação do vetor de entrada x(n) na camada de entrada. São calculados o campo local induzido e o sinal de saída de cada neurônio até a camada de saída, onde se obtém os resultados da rede y(n). Após calcula-se o erro de cada neurônio da camada de saída, pela comparação dos valores obtidos y (n) com os valores desejados d(n). Em seguida são calculados os erros globais instantâneos e médios para finalização.

Na quarta etapa, os sinais de erros são retro propagados, primeiramente calculando-se os gradientes locais para os neurônios da camada de saída e após para os neurônios da camada oculta. Em seguida, calculam-se os valores de ajuste para os pesos desta camada e os valores de bias, que como no cálculo da camada de saída devem ser somados aos valores atuais. Caso hajam mais camadas ocultas estas tarefas são efetuadas de forma idêntica até a camada de entrada onde os valores de ajuste da primeira camada oculta após a camada de entrada deve ter o valor $y_k(n)$ substituído pelo valor de $x_k(n)$.

Na última etapa, iteram-se todos os exemplos dos ciclos apresentados para treinamento da rede de forma aleatória até satisfazer o critério de parada. Este critério pode ter um número máximo de iterações ou um valor limite a ser atingido para o erro global médio da rede.

### 3.1.4 Aplicações e Exemplos

Entre as aplicações usuais das *RNA* têm-se: reconhecimento e classificação de padrões, *clustering*, previsão de séries temporais, aproximação de funções, predição,

otimização, setor militar (processamento de sinais para identificação de alvos e análise de imagens), sistemas especialistas, processamento de sinais (imagens, sensores, voz, caracteres, visão, compressão de dados, filtragem de sinais), telecomunicações, manufatura, monitoramento de processos e robótica ( CARVALHO, 2005 ; COELHO; JUNIOR, 2000).

Existem diferentes tipos de redes neurais artificiais, sendo o modelo de retro propagação do erro (LAWRENCE (1997), apud MELLO, 2004), o modelo mais estudado e aplicado para análises de mercado de ações. Este algoritmo, de aprendizado supervisionado, tem como objetivo, através do método do gradiente descendente, minimizar a função de erro entre a saída gerada pela rede neural e a saída desejada (GOLDSCHMIDT; PASSOS, 2005).

Como citado no início deste capítulo, diversos trabalhos nesta área foram desenvolvidos, alguns inclusive com resultados expressivos. Estes fatos motivam as pesquisas de modelos preditivos para previsão de ativos em bolsas de valores, aperfeiçoando ainda mais os conhecimentos até o momento obtidos.

## 3.2 Árvores de Decisão

A tarefa de classificação pode ser executada pela implementação de árvores de decisão e indução neural, ambas as técnicas baseadas no treinamento supervisionado. A indução neural é baseada na aplicação de redes neurais. Já as árvores de indução *(discovering)*, consistem em um método de representação de um modelo de padrões gerados a partir dos registros de treinamento. As árvores de indução podem ser representadas por modelos como árvores de decisão ou Regras “Se Então”. Como neste trabalho abordaremos somente as árvores de decisão, estas serão estudadas de forma detalhada a seguir. Alguns algoritmos baseados em árvores de decisão, são o ID3, o C4.5, o algoritmo CART e FID3.1.

O algoritmo C4.5 é uma evolução do algoritmo ID3, desenvolvido por Ross Quinlan em 1983 (PUC, 2004). O ID3 utiliza a entropia para avaliação das informações nos atributos, sendo que o atributo mais importante é colocado na raiz da árvore de decisão. Este algoritmo sofreu melhorias para corrigir uma limitação que não permitia ao ID3 trabalhar com atributos do tipo contínuo. Com isso, surgiu o C4.5, que além de aceitar valores contínuos, também

suporta valores desconhecidos, característica muito utilizada quando o modelo utilizado não possui padrões completos (PUC, 2004).

Árvores de decisão são métodos de mineração de dados, sendo geralmente construídos conforme a abordagem recursiva de particionamento de uma base de dados (GOLSCHMIDT; PASSOS, 2005). Devido a sua visualização, esta técnica facilita as decisões que envolvem riscos, pela forma organizada (gráfica) com que as variáveis relacionadas são apresentadas, e pode ser utilizada também para a realização de simulações. O resultado da simulação fornece uma base excelente para a tomada de decisão, pois mostra um conjunto de combinações de risco-retorno, ou seja, a relação entre os elementos que compõem a árvore (LACMAN, 1960 apud PEDRO; GUERREIRO, 2004).

As árvores de decisão são aplicadas em diversas áreas como ciências sociais, estatística, engenharia e inteligência artificial, diagnósticos médicos e avaliações de crédito para empréstimos. Árvores de decisão usadas para problemas de classificação são também chamadas de Árvores de Classificação (PUC, 2004).

De acordo com Goldschmidt e Passos (2005), a árvore de decisão é um modelo composto por nós, que representam decisões, e por ramos, que representam as alternativas possíveis para a decisão pontual. No final dos vários ramos existem folhas (nós que não possuem nós descendentes) significando os resultados que o conjunto de decisões anteriormente tomadas podem levar, estando associados a um rótulo ou um valor pela elevada homogeneidade dos elementos deste grupo (PUC, 2004). A cada nível da árvore é preciso definir regras heurísticas para separar os dados apresentados a este nó em subconjuntos homogêneos. (CARVALHO, 2005)

O funcionamento de uma árvore de decisão consiste na apresentação de um conjunto de dados ao nó inicial (raiz) da árvore. Conforme o resultado da decisão do cálculo de entropia dos diversos valores do conjunto de dados do nó inicial, a tupla, que é o conjunto de dados apresentado para a raiz da árvore (PUC, 2004), ramifica-se para um dos nós filhos. O objetivo deste processo é separar as classes de forma que tuplas de classes distintas tendam a ser associadas a diferentes partições. Este procedimento é repetido até que um nó terminal (folha) seja alcançado. Este procedimento caracteriza a recursividade da árvore de decisão.

As principais vantagens das árvores de decisão são sua eficiência computacional e simplicidade. Entretanto, o fato de alguns atributos aparecerem mais de uma vez na árvore de decisão pode gerar regras com informações relevantes, pois estes podem ser incluídos em todas as regras descobertas. Isso é considerado como desvantagem deste modelo, além disso, o processamento seria desperdiçado (AURÉLIO, VELLASCO; LOPES, 1999).

Neste algoritmo, o problema original é dividido em partes menores, semelhantes ao original e as soluções formarão uma combinação para o problema inicial (GONCHOROSKI, 2007). Para exemplificar a técnica de árvore de decisão, tomamos como base a tabela abaixo, que traz os dados de condições climáticas como aparência, temperatura e vento com objetivo de avaliar se o clima está apropriado para a prática de escaladas ou não.

**Tabela 3.0:** Entrada de dados para tarefa de classificação.

| Aparência | Temperatura | Vento | Escalar? |
|---|---|---|---|
| Sol | Quente | Não | Sim |
| Sol | Quente | Sim | Sim |
| Encoberto | Quente | Não | Não |
| Chuvoso | Agradável | Não | Não |
| Chuvoso | Frio | Não | Não |
| Chuvoso | Frio | Sim | Não |
| Encoberto | Frio | Sim | Não |
| Sol | Agradável | Não | Sim |
| Sol | Frio | Não | Sim |
| Chuvoso | Agradável | Não | Não |
| Sol | Agradável | Sim | Sim |
| Encoberto | Agradável | Sim | Sim |
| Encoberto | Quente | Não | Não |
| Chuvoso | Agradável | Sim | Não |

**Fonte:** Elaborado pelo autor.

O conhecimento descoberto é frequentemente representado na forma de regras **SE-ENTÃO**. Essas regras interpretam os atributos preditivos da tupla quanto à satisfação da condição antecedente da regra: "**SE** os atributos preditivos satisfazem as condições do antecedente da regra, **ENTÃO** a tupla tem a classe indicada no conseqüente da regra". Uma forma de realizar esta classificação é pela aplicação de modelos de árvores de decisão.

A partir de regras heurísticas no modelo de árvores de decisão, o resultado é uma estrutura como uma árvore, onde cada caminho é convertido em uma regra. O nó interno e os atributos das setas são tratados como antecedentes da regra (SE). Os atributos das folhas são a parte conseqüente das regras (Parte ENTÃO). Cabe salientar que um atributo pode ser

utilizado mais de uma vez na árvore de decisão, assim como podem existir ramificações onde certos atributos não são avaliados (PUC, 2004). Para se obter a recursividade explicada anteriormente, cada tupla tem seu valor testado já no nó raiz e, após, é testada nos nós inferiores, seguindo as setas, em função do valor do atributo testado no nó em questão. Este processo é repetido até que a tupla chegue na folha. No exemplo citado (tabela 3.0), os campos Aparência, Temperatura e Vento são utilizados para responder ao questionamento sobre a condição para a prática de escaladas ou não. A seguir é apresentado um dos principais algoritmos baseado em árvores de decisão, o C4.5.

### 3.2.1 Algoritmo C4.5

Este algoritmo tem como objetivo abstrair árvores de decisão baseado em um método recursivo de particionamento das bases de dados. A maioria destes algoritmos consiste em duas fases, a construção da árvore de decisão e sua simplificação.

Após a geração da árvore, o conjunto de treinamento é separado em duas ou mais partições de acordo com os valores do atributo alvo de cada nó. Este processo é repetido até que todos, ou a maioria dos exemplos pertençam a uma classe. As árvores são divididas em níveis, de forma que todos os níveis devem ser processados antes do processamento do nível seguinte.

Os passos deste algoritmo consistem em primeiramente, escolher um atributo. Após, adicionar um ramo para cada atributo identificado. Na seqüência, conforme o valor do atributo escolhido, o exemplo ou tupla é passado para os respectivos nós filhos, até alcançar alguma folha. Por último, deve-se associar a cada folha uma classe, senão repetir os passos anteriores.

O algoritmo realiza a poda com base nos erros de seus nós e descendentes, baseados nos cálculos de entropia e do ganho de informação. Entropia é a medida da relação de informações no sistema. Quanto maior a entropia, menor a quantidade de informações deste sistema. O ganho de informação, conforme (GONCHOROSKI, 2007) mede a redução de

entropia nas partições dos exemplos de treinamento, conforme o valor do atributo. Com isso, são geradas árvores menores e menos complexas.

Em um dado nó da árvore e com as tuplas do conjunto de treinamento S, é selecionado o melhor atributo para o dado nó. Para cada valor (vi) deste atributo, deve ser criada uma sub-árvore ou uma folha sob este nó. Para definição do melhor atributo, existe o seguinte critério:

Sejam S todas as tuplas do conjunto de treinamento, A um atributo, s uma tupla, v um valor, e c o número de classes, define-se:

**Equação 1**: $S_v = \{s \in S \mid A(s) = v\}$

A Equação 1 representa as tuplas do conjunto de treinamento que no atributo A possuem o valor v. Com isso, é definida a entropia do conjunto S, como:

**Equação 2:** $\text{Entropia}(S) = \sum i=1, ..., c \; -pi * \log(pi)$

Onde pi é a probabilidade de ocorrência de uma determinada classe e, com isso, pode-se definir o ganho da escolha do atributo A com relação ao conjunto de treinamento (S) conforme a equação a seguir:

**Equação 3:** $\text{Ganho}(S, A) = \text{Entropia}(S) - \sum v \in \text{valores}(A) \; (|Sv| / |S|) * \text{Entropia}(Sv)$

O ganho de informação mede a eficácia de um atributo nos dados de treinamento. A escolha do atributo de menor entropia faz com que sejam geradas árvores com menos nós e ramificações. A seguir é apresentado um pseudocódigo do algoritmo C4.5.

```
Função C4.5
  (R:Conjunto de atributos não classificadores,
   C Atributo classificador,
   S Conjunto de treinamento ) retorna árvore de decisão;
Início:
  Se S está vazio,
        retornar um único nó com valor de FALHA;
  Se todos os registros de S tem o mesmo valor para o atributo classificador,
```

```
        retornar um único nó com este valor;
  Se R está vazio,
        retornar um único nó com o valor mais freqüente do atributo classificador do conjunto S;
  Se R não está vazio,
        D = atributo com maior valor de ganho (D,S) dos atributos de R;
        Sejam {dj |j= 1,2,...m} os valores do atributo D;
        Sejam {dj |j=1, 2,...m} os subconjuntos de S correspondentes aos valores de dj respectivamente;
        Devolver uma árvore com a raiz nomeada como D, com arcos nomeados (d1,d2,...dm), que vão
respectivamente às árvores: C4.5(R-{D}, C, S1), C4.5(R-{D}, C, S2), C4.5(R-{D}, C, Sm)
Fim.
```

**Quadro 3.1 –** Pseudocódigo Algoritmo C4.5
**Fonte:** GONCHOROSKI, 2007.

No caso do exemplo da tabela 3.0, as possibilidades para o atributo raiz são:

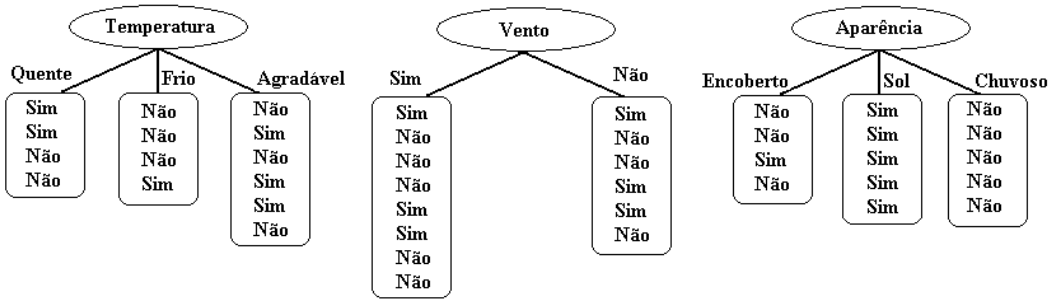


**FIGURA 3.4:** Atributos para a raiz da árvore de decisão.
**Fonte:Do autor**

Para identificar a raiz adequada ao modelo apresentado, deve ser calculada a entropia para cada um dos atributos possíveis para a raiz e seus campos. Feito isso, deve-se calcular o ganho de cada atributo, onde o atributo adequado irá apresentar o maior ganho, conforme o algoritmo C4.5, gerando a menor árvore.

Identificadas as possíveis árvores, é selecionada uma como modelo e aplicada ao conjunto de testes. Esta técnica de fácil visualização, utilizada para classificação de itens conforme os valores de seus atributos poderá ser aplicada no mercado financeiro, classificando os ativos de acordo com os valores de cotação do preço das ações, valores de abertura e fechamento, volumes negociados, etc. Com isso, uma análise indicará ao investidor as possibilidades de ganhos com o ativo analisado, de acordo com as sazonalidades e padrões identificados na amostra.

No capítulo seguinte é abordado o estudo de caso na Bovespa, demonstrando como foram realizadas as tarefas da etapa de pré-processamento da base de dados.

# 4 ESTUDO DE CASO NA BOVESPA

Como os dados a serem analisados neste trabalho são dados de domínio público, estes podem ser obtidos por alguns meios eletrônicos, como o site da Bovespa, compra de empresas que vendem estes dados, através de sites de cotações de ações da Bovespa, como o ADVFN[3], ou ainda através de programas proprietários de análise de mercados, como o CMA[4].

Neste trabalho, primeiramente foi efetuada a busca dos dados no site da instituição, no link: <http://www.bovespa.com.br/Mercado/RendaVariavel/SeriesHistoricas/FormSeriesHistoricas.asp> (Bovespa, 2007), onde é possível, após um cadastro, obter os dados históricos diários dos preços dos títulos negociados na Bolsa desde 1986. As cotações são fornecidas na moeda e forma de cotação da época, sem nenhum ajuste para a inflação ou proventos, por exemplo. Também é possível a obtenção de dados a cada minuto, mas somente dos últimos 12 meses.

Estes arquivos, depois de realizado o download e descompactação, possuem as principais informações dos ativos, como: nome e código da empresa, código da ação, código ISIN[5], tipo de mercado (á vista, termo, opções), especificação (ON/PN), preços (anterior, abertura, mínimo, médio, máximo, fechamento), quantidade de negócios e volume negociado.

---

[3] www.advfn.com.br

[4] Encontrado em www.cma.com.br

[5] Código ISIN é a identificação do título conforme a Norma ISO 6166, que é uma padronização internacional para identificação de títulos financeiros.

As informações, estão em um formato de arquivo texto (.txt) sem identificador de colunas ou marcadores e com o dados de todos os pregões diários de todos os ativos da instituição sem classificação.

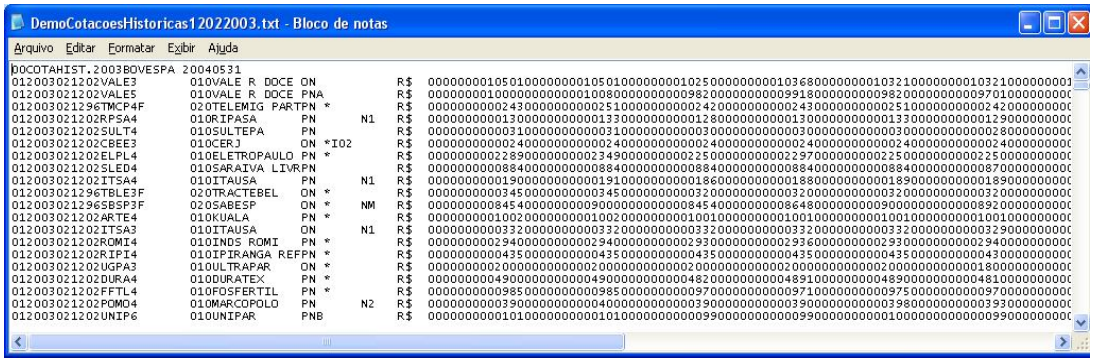


**Figura 4.1: Exemplo de *layout* de arquivo diário**
**Fonte: Do autor**

O *layout* deste arquivo (figura 4.1), devido a complexidade deve ser interpretada conforme orientações da instituição, disponíveis em < http://www.bovespa.com.br/Pdf/SeriesHistoricas_Layout.pdf>.

Neste arquivo, informações como o formato do nome do arquivo, que identifica o ano correspondente ( EX: COTAHIST.2003.TXT ) e estrutura do arquivo, com informações sobre nome de cada campo, descrição do conteúdo, tipo e tamanho de cada registro e a posição inicial e final de cada registro são disponibilizadas.

Para importação deste arquivo para o Excel (formato de tabelas), foram seguidos os passos conforme o tutorial disponibilizado em <http://www.bovespa.com.br/Home/redirect.asp?end=/Mercado/RendaVariavel/SeriesHistoricas/FormTutorial.asp>, onde passo-a-passo a conversão do arquivo texto é realizada, utilizando os *layouts* explicados anteriormente. O formato final deste arquivo é demonstrado na figura 4.2.

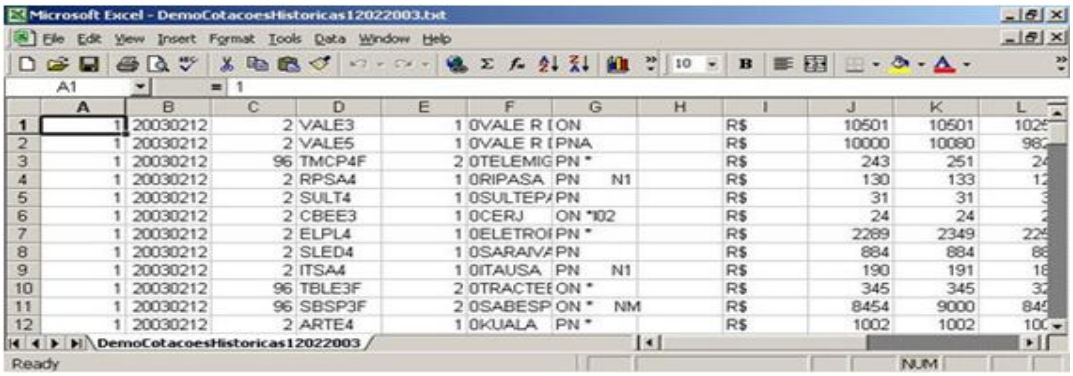


**Figura 4.2:** Arquivo diário separado por colunas.
**Fonte:** Do autor.

Com este arquivo já configurado, e com o código do ativo a ser estudado, basta realizar um filtro pelo código e separar os dados diários do respectivo ativo, no ano abordado. Este processo de preparação de dados, com a utilização do software proprietário é desnecessário.

A opção de comprar os dados de empresas que fornecem informações macroeconômicas de mercados financeiros e de setores da economia foi abordada. Foram estudadas duas empresas, para um orçamento de dados diários no período de 2003 a 2008, sendo elas a QuoteBR, encontrada em < www.**quotebr**.com > e Lafis, encontrada em < http://www.lafis.com.br>. Obtivemos resposta de uma das empresas, que orçou os dados em R$ 1200,00 reais, sendo que 2008 seriam dados até abril.

O site de cotações de ações da Bovespa e demais mercados, ADVFN, encontrado em < http://br.advfn.com> disponibiliza o download de dados históricos da Bovespa, selecionados pelo código do ativo, diário, desde 1991 e *intraday*, a cada minuto, mas somente das últimas duas semanas a contar da data de consulta. Como inicialmente o trabalho necessitava de um histórico de anos com o intervalo a cada 10 ou 15 minutos, a opção foi descartada.

A utilização de um software proprietário para a obtenção dos dados se tornou a principal opção por fornecer para download os dados a cada 15 minutos e os dados diários dos ativos selecionados. Outro fator que contou para a escolha desta opção foi a disponibilidade de um investidor pessoa física conhecido, que disponibilizou a ferramenta para a importação

dos dados e para uso do sistema quando necessário. O arquivo é exportado já em um formato CSV, como mostrado na figura 4,3.

| [012] PETR4 (D) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data | Abert | Máxima | Mínima | Fech | Qtde papel | BB1 (8) (F | BB2 (8) (F | BB3 (8) (F | M1_M2 (3 | M2_M2 (3 | M3_M2 (3 | IFR (7) (F) | VOL |
| 26/3/2003 | 4,58 | 4,6 | 4,53 | 4,53 | 7357600 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7357600 |
| 27/3/2003 | 4,49 | 4,54 | 4,43 | 4,54 | 9105600 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9105600 |
| 28/3/2003 | 4,6 | 4,68 | 4,54 | 4,6 | 6137600 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6137600 |
| 31/3/2003 | 4,56 | 4,58 | 4,54 | 4,58 | 4844800 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4844800 |
| 1/4/2003 | 4,61 | 4,74 | 4,61 | 4,71 | 11840000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11840000 |
| 2/4/2003 | 4,74 | 4,82 | 4,74 | 4,78 | 11483200 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11483200 |
| 3/4/2003 | 4,8 | 4,9 | 4,78 | 4,84 | 8910400 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8910400 |
| 4/4/2003 | 4,85 | 4,88 | 4,78 | 4,86 | 8884800 | 4,68 | 4,4102 | 4,9498 | 0 | 0 | 0 | 94,5946 | 8884800 |
| 7/4/2003 | 4,94 | 5 | 4,72 | 4,72 | 13140000 | 4,7037 | 4,4623 | 4,9452 | 0 | 0 | 0 | 68 | 13140000 |
| 8/4/2003 | 4,74 | 4,76 | 4,54 | 4,54 | 11715200 | 4,7037 | 4,4623 | 4,9452 | 0 | 0 | 0 | 45,1613 | 11715200 |
| 9/4/2003 | 4,57 | 4,6 | 4,51 | 4,57 | 7840000 | 4,7 | 4,4504 | 4,9496 | 0 | 0 | 0 | 49,2063 | 7840000 |

**Figura 4.3:** Exemplo de arquivo diário exportado
**Fonte:** Do autor

## 4.1 Os Dados Estudados

Como foi visto anteriormente, a aplicação das técnicas de *data mining* no mercado financeiro, assim como em qualquer outro contexto, precisa seguir as etapas do processo de descoberta de conhecimento, citadas no capítulo 2.

### 4.1.1 Seleção

Conforme Rezende (2005), seleção é a fase onde os dados realmente úteis para o objetivo do estudo são identificados. Para tanto, a base de dados a ser utilizada precisa ser analisada, identificando quais os atributos que serão estudados e após minerados, conforme Goldschmidt e Passos (2005).

Com este objetivo, primeiramente deve ser realizada a seleção dos dados. No ambiente abordado, precisam ser identificados dentro do mercado à vista da Bovespa os ativos que farão parte deste grupo. Como regra para escolha do ativo a ser abordado, foi utilizado como critério de seleção a sua participação dentro do índice Ibovespa durante os anos de 2006 e 2007 (BOVESPA, 2007). Este levantamento é realizado pela Bovespa a cada quadrimestre. Na tabela a seguir estão identificados os 10 maiores contribuintes no índice

Ibovespa e seu percentual de participação. Como a ação preferencial da Petrobrás (PETR4) destaca-se neste estudo, foi selecionada para ser utilizada como estudo de caso na aplicação das técnicas de *Data Mining*.

**Tabela 4.0:** Percentual de participação dos 10 maiores contribuintes do Ibovespa para cada quadrimestre:

| Ano => | | 2006 | | | | | | 2007 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CÓD. | AÇÃO | Jan. - Abr | | Mai. - Ago. | | Set. - Dez | | Jan. - Abr | | Mai. - Ago. | | Set. - Dez. | |
| | | % | Pos | % | Pos | % | Pos | % | Pos | % | Pos | % | Pos |
| PETR4 | PETROBRAS | 9,227 | 1 | 11,278 | 1 | 13,086 | 1 | 13,798 | 1 | 13,85 | 1 | 13,689 | 1 |
| TNLP4 | TELEMAR | 8,117 | 2 | 6,567 | 3 | 4,724 | 4 | 3,581 | 5 | 3,037 | 6 | 2,289 | 9 |
| VALE5 | VALE R DOCE | 8,095 | 3 | 8,401 | 2 | 11,058 | 2 | 9,955 | 2 | 9,787 | 2 | 10,426 | 2 |
| USIM5 | USIMINAS | 5,593 | 4 | 5,215 | 4 | 4,49 | 5 | 4,121 | 4 | 3,999 | 4 | 3,412 | 4 |
| CSNA3 | SID NACIONAL | 4,23 | 5 | 3,66 | 6 | 2,985 | 7 | 2,523 | 8 | 2,339 | 10 | 2,285 | 10 |
| CMET4 | CAEMI | 4,128 | 6 | 3,617 | 7 | - | | - | | - | | - | |
| BBDC4 | BRADESCO | 3,777 | 7 | 4,258 | 5 | 4,815 | 3 | 4,537 | 3 | 4,138 | 3 | 3,929 | 3 |
| GGBR4 | GERDAU | 3,512 | 8 | 3,161 | 8 | 2,883 | 8 | 2,653 | 7 | 2,632 | 8 | 2,544 | 7 |
| BRKM5 | BRASKEM | 2,914 | 9 | 2,935 | 9 | 2,417 | 10 | - | | - | | - | |
| ITAU4 | ITAUBANCO | 2,77 | 10 | 2,83 | 10 | 3,455 | 6 | 3,312 | 6 | 3,158 | 5 | 2,859 | 5 |
| VALE3 | VALE R DOCE | - | | - | | 2,517 | 9 | 2,447 | 10 | 2,647 | 7 | 2,72 | 6 |
| PETR3 | PETROBRÁS | - | | - | | - | | 2,473 | 9 | 2,436 | 9 | 2,461 | 8 |

**Fonte:** Do autor

A idéia inicial do trabalho foi a aplicação das técnicas de *Data Mining* em um período de amostragem diário e em um período de amostragem a cada 15 minutos. Entretanto, para delimitação do escopo e foco dos esforços, optou-se por trabalhar com os dados diários, visto que se forem descobertas relações entre os atributos estudados, a mesma técnica pode ser utilizada após para descoberta de padrões nos períodos de 15 minutos.

Logo, o objeto de estudo alvo deste trabalho é a identificação de padrões no período amostrado de 1/04/2003 até 3/7/2008 para atributos utilizados na análise técnica de ativos, em um período de amostragem diário. As amostras consistem em indicações de preço (cotação) de abertura e fechamento para o dia de cada ação, volume de negociações e o IFR, sendo este um estudo que relaciona o preço e os movimentos de compra e vendas do ativo no mercado.

Como muitos investidores utilizam o dólar como parâmetro para negociações na bolsa e o produto alvo da companhia estudada é negociado em dólar, julgou-se necessário o acompanhamento do desempenho da cotação desta moeda juntamente com os atributos do ativo.

Outro atributo a ser levado em consideração é a movimentação diária do índice Dow Jones e seu volume de negociações, visto que a empresa tem grande parte de seu patrimônio também negociado na bolsa de valores norte-americana.

Por fim, como o ativo representa, conforme a tabela anteriormente citada, mais de 10% do volume de negociações da Bovespa, a cotação do índice da Bovespa e seu volume de negociações também tornam-se atributos que podem interferir na cotação deste ativo, e assim, tornam-se alvos deste estudo.

Um atributo que seria interessantíssimo para este estudo seria o preço do barril de petróleo cru no mercado internacional. Entretanto, devido a impossibilidade de acesso a cotação histórica deste ativo, ele não foi abordado neste trabalho. Porém cabe salientar que a falta desta informação pode influenciar nos resultados obtidos.

### 4.1.2 Limpeza

Os dados foram exportados à partir do software proprietário CMA, em formato "csv". Cada arquivo, de cada ativo, PETR4, IBOV, Dow Jones e dólar, trazia as informações de estudos do usuário do sistema. Desta forma, os dados precisavam ser classificados, deixando em cada arquivo somente os atributos realmente úteis a este estudo. Neste ponto do trabalho também foi definido como referência entre os ativos os valores de abertura, volumes e fechamento diários, pois o objetivo é a classificação do valor futuro de fechamento do ativo estudado, no caso PETR4.

Como exemplo, é mostrado na figura 4.4 o arquivo referente ao ativo da PETR4. Os dados originais exportados eram constituídos de, conforme data, valor de abertura, máxima,mínima, fechamento, Quantidade de papel (títulos ou unidades negociadas) (BB1 (8) (F), BB2 (8) (F), BB3 (8) (F), M1_M2 (3, 8, 20) (F), M2_M2 (3, 8, 20) (F), M3_M2 (3, 8, 20) (F), IFR (7) (F)) e Volume. Os registros de (BB1 (8) (F), BB2 (8) (F), BB3 (8) (F), M1_M2 (3, 8, 20) (F), M2_M2 (3, 8, 20) (F), M3_M2 (3, 8, 20) (F)) são referentes aos estudos aplicados pelo analista técnico no gráfico, e que como os dados de mínimo, máximo e quantidade de papel negociado deveriam ser expurgados.

| [012] PETR4 (D) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data | Abert | Máxima | Mínima | Fech | Qtde pap | BB1 (8) (F | BB2 (8) (F | BB3 (8) (F | M1_M2 (3 | M2_M2 (3 | M3_M2 (3 | IFR (7) (F | VOL |
| 26/3/2003 | 4,58 | 4,6 | 4,53 | 4,53 | 7357600 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7357600 |
| 27/3/2003 | 4,49 | 4,54 | 4,43 | 4,54 | 9105600 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9105600 |
| 28/3/2003 | 4,6 | 4,68 | 4,54 | 4,6 | 6137600 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6137600 |
| 31/3/2003 | 4,56 | 4,58 | 4,54 | 4,58 | 4844800 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4844800 |
| 1/4/2003 | 4,61 | 4,74 | 4,61 | 4,71 | 1,2E+07 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1,2E+07 |
| 2/4/2003 | 4,74 | 4,82 | 4,74 | 4,78 | 1,1E+07 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1,1E+07 |
| 3/4/2003 | 4,8 | 4,9 | 4,78 | 4,84 | 8910400 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8910400 |
| 4/4/2003 | 4,85 | 4,88 | 4,78 | 4,86 | 8884800 | 4,68 | 4,4102 | 4,9498 | 0 | 0 | 0 | 94,5946 | 8884800 |
| 7/4/2003 | 4,94 | 5 | 4,72 | 4,72 | 1,3E+07 | 4,7037 | 4,4623 | 4,9452 | 0 | 0 | 0 | 68 | 1,3E+07 |
| 8/4/2003 | 4,74 | 4,76 | 4,54 | 4,54 | 1,2E+07 | 4,7037 | 4,4623 | 4,9452 | 0 | 0 | 0 | 45,1613 | 1,2E+07 |
| 9/4/2003 | 4,57 | 4,6 | 4,51 | 4,57 | 7840000 | 4,7 | 4,4504 | 4,9496 | 0 | 0 | 0 | 49,2063 | 7840000 |
| 10/4/2003 | 4,58 | 4,65 | 4,57 | 4,57 | 1,1E+07 | 4,6987 | 4,4463 | 4,9512 | 0 | 0 | 0 | 36 | 1,1E+07 |
| 11/4/2003 | 4,58 | 4,63 | 4,56 | 4,58 | 6458400 | 4,6825 | 4,417 | 4,948 | 0 | 0 | 0 | 27,2727 | 6458400 |
| 14/4/2003 | 4,6 | 4,65 | 4,57 | 4,58 | 6769600 | 4,6575 | 4,3963 | 4,9187 | 0 | 0 | 0 | 15,7895 | 6769600 |
| 15/4/2003 | 4,6 | 4,62 | 4,54 | 4,62 | 5275200 | 4,63 | 4,4143 | 4,8457 | 0 | 0 | 0 | 20 | 5275200 |
| 16/4/2003 | 4,64 | 4,68 | 4,58 | 4,58 | 9974400 | 4,595 | 4,4849 | 4,7051 | 0 | 0 | 0 | 26,6667 | 9974400 |
| 17/4/2003 | 4,65 | 4,72 | 4,63 | 4,7 | 6743200 | 4,5925 | 4,4953 | 4,6897 | 0 | 0 | 0 | 83,3333 | 6743200 |
| 22/4/2003 | 4,65 | 4,76 | 4,64 | 4,76 | 9965600 | 4,62 | 4,477 | 4,763 | 0 | 0 | 0 | 85,1852 | 9965600 |
| 23/4/2003 | 4,74 | 4,83 | 4,7 | 4,81 | 9642400 | 4,65 | 4,4615 | 4,8385 | 0 | 0 | 0 | 87,5 | 9642400 |
| 24/4/2003 | 4,8 | 4,85 | 4,72 | 4,74 | 7388000 | 4,6712 | 4,4856 | 4,8569 | 1,0211 | 1 | 0,9977 | 71,0526 | 7388000 |
| 25/4/2003 | 4,74 | 4,77 | 4,69 | 4,69 | 6731200 | 4,685 | 4,5146 | 4,8554 | 1,0132 | 1 | 0,9965 | 62,7907 | 6731200 |

**Figura 4.4:** Atributos de ativo da Bovespa
**Fonte:** Do autor

A limpeza de atributos foi efetuada em todos ativos abordados, PETR4, IBOV, dólar e Dow Jones. Feito isso, os dados, ainda com as datas originais, foram reunidos em uma única base (Figura 4.5).

| [012] PETR4 (D) | | | | | [012] IBOV (D) | | | [096] DJI (D) | | | [062] DOL COM (D) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data | Abert | Fech | IFR (7) (F) | VOL | Data | Fech | VOL | Data | Fech | VOL | Data | Abert |
| 26/11/2003 | 6,96 | 6,83 | 75,7576 | 12071200 | 26/11/2003 | 19694 | 125026753 | 26/11/2003 | 9779,57 | 162206433 | 25/11/2003 | 2,924 |
| 28/11/2003 | 6,91 | 6,92 | 70,3704 | 6196800 | 28/11/2003 | 20183 | 106851245 | 28/11/2003 | 9782,46 | 79157359 | 26/11/2003 | 2,936 |
| 1/12/2003 | 6,94 | 7,04 | 78,9474 | 5957600 | 1/12/2003 | 20520 | 155119416 | 1/12/2003 | 9899,05 | 227651855 | 27/11/2003 | 2,952 |
| 2/12/2003 | 7,02 | 7,1 | 85,7143 | 8302400 | 2/12/2003 | 20458 | 118197810 | 2/12/2003 | 9853,64 | 218174995 | 28/11/2003 | 2,961 |
| 3/12/2003 | 7,1 | 7,02 | 72 | 7774400 | 3/12/2003 | 20539 | 127415998 | 3/12/2003 | 9873,42 | 222568837 | 1/12/2003 | 2,95 |
| 4/12/2003 | 7,02 | 7,06 | 68,8889 | 8096000 | 4/12/2003 | 20414 | 132060745 | 4/12/2003 | 9930,82 | 216936236 | 2/12/2003 | 2,935 |
| 5/12/2003 | 7,04 | 7,1 | 81,3953 | 9979200 | 5/12/2003 | 20879 | 133847703 | 5/12/2003 | 9862,68 | 201637332 | 3/12/2003 | 2,942 |
| 8/12/2003 | 7,08 | 7,11 | 78,3784 | 7419200 | 8/12/2003 | 20888 | 108897857 | 8/12/2003 | 9965,27 | 192729102 | 4/12/2003 | 2,939 |
| 9/12/2003 | 7,14 | 7,38 | 87,0968 | 20095200 | 9/12/2003 | 21259 | 183434767 | 9/12/2003 | 9923,42 | 231393377 | 5/12/2003 | 2,954 |
| 10/12/2003 | 7,36 | 7,36 | 80,7692 | 13102400 | 10/12/2003 | 20972 | 157213477 | 10/12/2003 | 9921,86 | 220980731 | 8/12/2003 | 2,945 |
| 11/12/2003 | 7,38 | 7,56 | 84,8485 | 10138400 | 11/12/2003 | 21296 | 121258304 | 11/12/2003 | 10008,16 | 208632426 | 9/12/2003 | 2,948 |
| 12/12/2003 | 7,6 | 7,48 | 84,8485 | 8378400 | 12/12/2003 | 20973 | 86354401 | 12/12/2003 | 10042,16 | 176896872 | 10/12/2003 | 2,948 |
| 15/12/2003 | 7,55 | 7,42 | 76,4706 | 11815200 | 15/12/2003 | 20709 | 131117576 | 15/12/2003 | 10022,82 | 247017244 | 11/12/2003 | 2,954 |
| 16/12/2003 | 7,43 | 7,49 | 77,4648 | 6960000 | 16/12/2003 | 20759 | 111390100 | 16/12/2003 | 10129,56 | 233731105 | 12/12/2003 | 2,949 |
| 17/12/2003 | 7,5 | 7,66 | 81,6092 | 10035200 | 17/12/2003 | 21199 | 113007678 | 17/12/2003 | 10145,26 | 197177001 | 15/12/2003 | 2,944 |
| 18/12/2003 | 7,64 | 7,8 | 78,3784 | 7045600 | 18/12/2003 | 21489 | 85744700 | 18/12/2003 | 10248,08 | 218014239 | 16/12/2003 | 2,932 |
| 19/12/2003 | 7,76 | 7,84 | 81,5789 | 7818400 | 19/12/2003 | 21385 | 83455301 | 19/12/2003 | 10278,22 | 279780355 | 17/12/2003 | 2,95 |
| 22/12/2003 | 7,84 | 8,02 | 81,0811 | 6371200 | 22/12/2003 | 21630 | 57038396 | 22/12/2003 | 10338 | 165729938 | 18/12/2003 | 2,936 |
| 23/12/2003 | 8,02 | 7,86 | 73,1707 | 6741600 | 23/12/2003 | 21688 | 91751906 | 23/12/2003 | 10341,26 | 151085076 | 19/12/2003 | 2,938 |
| 26/12/2003 | 7,87 | 7,74 | 68,1818 | 4528000 | 26/12/2003 | 21806 | 29973499 | 26/12/2003 | 10324,67 | 49518954 | 22/12/2003 | 2,934 |
| 29/12/2003 | 7,78 | 7,92 | 71,7172 | 7999200 | 29/12/2003 | 22045 | 51566826 | 29/12/2003 | 10450 | 156396552 | 23/12/2003 | 2,929 |
| 30/12/2003 | 7,96 | 7,92 | 65,8537 | 5283200 | 30/12/2003 | 22236 | 56869761 | 30/12/2003 | 10425,04 | 132795236 | 24/12/2003 | 2,912 |
| 2/1/2004 | 7,92 | 8,2 | 70,8333 | 5923200 | 2/1/2004 | 22444 | 33341957 | 2/1/2004 | 10409,85 | 168890195 | 26/12/2003 | 2,916 |
| 5/1/2004 | 8,26 | 8,67 | 79,8561 | 11188000 | 5/1/2004 | 23531 | 137081482 | 5/1/2004 | 10544,07 | 221288980 | 29/12/2003 | 2,907 |
| 6/1/2004 | 8,79 | 9,12 | 83,1325 | 19296800 | 6/1/2004 | 23576 | 138875283 | 6/1/2004 | 10538,66 | 191462498 | 30/12/2003 | 2,869 |
| 7/1/2004 | 9,1 | 9,02 | 86,25 | 17758400 | 7/1/2004 | 23320 | 162836784 | 7/1/2004 | 10529,03 | 225486973 | 2/1/2004 | 2,9 |
| 8/1/2004 | 9,21 | 8,97 | 90,1961 | 13824000 | 8/1/2004 | 23716 | 141681792 | 8/1/2004 | 10592,44 | 237773419 | 5/1/2004 | 2,88 |

**Figura 4.5:** Dados dos ativos reunidos.
**Fonte:** Do autor

A base, com 1304 registros deve ainda sofrer outro processo de limpeza, desta vez, alinhando os dados diários das bases de dados com origem no Brasil e os dados com origem nos Estados Unidos, pois devido a feriados nacionais, recessos culturais como o carnaval no Brasil e outros fatores, há  falta de dados de ambas as partes, em datas diferentes, como na figura a seguir.

| [012] PETR4 (D) | | | | |
|---|---|---|---|---|
| Data | Abert | Fech | IFR (7) (F) | VOL |
| 26/6/2007 | 25 | 24,72 | 49,2958 | 11645400 |
| 27/6/2007 | 24,62 | 24,96 | 53,8462 | 15054800 |
| 28/6/2007 | 25,01 | 24,96 | 40,9836 | 13258000 |
| 29/6/2007 | 25,1 | 25,18 | 70,5882 | 13756600 |
| 2/7/2007 | 25,25 | 25,74 | 77,4436 | 12750200 |
| 3/7/2007 | 25,78 | 26,18 | 82,9545 | 14023400 |
| 4/7/2007 | 26,2 | 26,3 | 86,8132 | 8255000 |
| 5/7/2007 | 26,19 | 26,44 | 100 | 16351200 |
| 6/7/2007 | 26,5 | 26,68 | 100 | 12786200 |
| | | | | |
| 10/7/2007 | 26,69 | 26,25 | 80 | 18811600 |
| 11/7/2007 | 26,32 | 26,46 | 79,9065 | 13848400 |
| 12/7/2007 | 26,64 | 27,05 | 80,1843 | 16427400 |
| 13/7/2007 | 26,93 | 27,22 | 77,3684 | 19117400 |
| 16/7/2007 | 27,22 | 26,96 | 66,1765 | 17728800 |
| 17/7/2007 | 26,86 | 27,44 | 71,0084 | 15330400 |

| [012] IBOV (D) | | |
|---|---|---|
| Data | Fech | VOL |
| 26/6/2007 | 53851 | 13671211 |
| 27/6/2007 | 54143 | 14941473 |
| 28/6/2007 | 54146 | 12678703 |
| 29/6/2007 | 54392 | 21109192 |
| 2/7/2007 | 55371 | 14973866 |
| 3/7/2007 | 55699 | 14275335 |
| 4/7/2007 | 55696 | 17084013 |
| 5/7/2007 | 55932 | 31610382 |
| 6/7/2007 | 56443 | 16448905 |
| | | |
| 10/7/2007 | 55882 | 16207389 |
| 11/7/2007 | 56356 | 38806958 |
| 12/7/2007 | 57613 | 24578565 |
| 13/7/2007 | 57644 | 12845336 |
| 16/7/2007 | 57374 | 12350890 |
| 17/7/2007 | 57659 | 13339672 |

| [096] DJI (D) | | |
|---|---|---|
| Data | Fech | VOL |
| 26/6/2007 | 13337,66 | 240948351 |
| 27/6/2007 | 13427,73 | 246019415 |
| 28/6/2007 | 13422,28 | 207127217 |
| 29/6/2007 | 13408,62 | 262107608 |
| 2/7/2007 | 13535,43 | 196411638 |
| 3/7/2007 | 13577,3 | 111584923 |
| | | |
| 5/7/2007 | 13565,84 | 188836639 |
| 6/7/2007 | 13611,68 | 176035608 |
| 9/7/2007 | 13649,97 | 192835577 |
| 10/7/2007 | 13501,7 | 274423909 |
| 11/7/2007 | 13577,87 | 224412879 |
| 12/7/2007 | 13861,73 | 300519805 |
| 13/7/2007 | 13907,25 | 223816717 |
| 16/7/2007 | 13950,98 | 209066573 |
| 17/7/2007 | 13971,55 | 266018610 |

| [062] DOL COM (D) | |
|---|---|
| Data | Abert |
| 26/6/2007 | 1,94 |
| 27/6/2007 | 1,966 |
| 28/6/2007 | 1,936 |
| 29/6/2007 | 1,927 |
| 2/7/2007 | 1,924 |
| 3/7/2007 | 1,91 |
| 4/7/2007 | 1,911 |
| 5/7/2007 | 1,909 |
| 6/7/2007 | 1,909 |
| 9/7/2007 | 1,908 |
| 10/7/2007 | 1,901 |
| 11/7/2007 | 1,903 |
| 12/7/2007 | 1,884 |
| 13/7/2007 | 1,865 |
| 16/7/2007 | 1,864 |
| 17/7/2007 | 1,864 |

**Figura 4.6:** Exemplos da falta registros nos ativos estudados.
**Fonte:** Do autor

As opções neste tipo de operação são o preenchimento do valor ausente como, por exemplo, a média de valores de um período definido, ou a exclusão dos itens deste registro presente no ativo vizinho. As duas opções podem causar interferências no resultado final da pesquisa, fortemente influenciada pelo número total de registros presentes na base utilizada. No caso abordado, ficou definida a exclusão dos valores presentes nos ativos vizinhos. Com isso, a base utilizada ficou com 1277 registros válidos para o processo de mineração de dados.

### 4.1.3 Tratamento

A base de dados já definida e filtrada, durante o pré-processamento ainda precisa passar por um processo conhecido como tratamento dos dados. Este processo visa adequar os dados para aplicação do algoritmo de mineração definido.

Inicialmente os dados foram padronizados com a diferença numérica entre o dia atual e o dia anterior, visando minimizar as diferenças entre os valores abordados nos atributos, pois os valores de preço de abertura, por exemplo, variavam de R$ 4,61 a R$ 45,00. Neste ponto surgiu um problema, após a confecção da base e tratamento dos valores. Como manter a mesma forma de avaliação para preços desde 2003 até o ano de 2008? Por exemplo: Em 2003 o preço da ação da PETR4 estava em R$ 4,00 por ação. Este mesmo dado terá que ser avaliado com o preço da mesma ação em 2008, já na faixa de R$ 44,00. De forma mais clara, em 2003 a variação de preço de um dia para outro de R$ 4,00 para R$ 4,40 resultava em uma diferença de R$ 0,40, que em termos percentuais corresponde a 10% do valor da ação. Já em

2008, a simples variação de preço de R$ 44,00 para R$ 44,40 é a mesma, R$ 0,40 por ação. Entretanto, em termos percentuais esta diferença corresponde a 1% do valor da ação.

Conforme definido pela própria instituição abordada, a variação do dia indica a variação do preço de um ativo, como uma ação, cota de fundo, ou valor de qualquer título, durante o dia. Caso o pregão ainda não tenha encerrado, reflete a variação acumulada no dia até aquele momento e é calculada com relação ao fechamento no dia anterior.

Na Bovespa, as variações de preço são tratadas em percentuais, com objetivo de ter uma forma de medição de desempenho e variação homogênea para todos os anos. Assim, neste estudo, o formato dos valores de todos os atributos, inclusive os volumes diários foram transformados para o formato percentual, sempre com relação ao dia anterior. Desta forma, a medição é homogênea em todas as épocas de amostragem. O cálculo segue o seguinte padrão: A variação diária seja ela de preço ou volume é calculada da seguinte forma:

*Variação de Hoje = (Valor de hoje / Valor de ontem)-1*.

Assim, todos os valores de todos os atributos são tratados em percentual, facilitando o processo de padronização dos dados.

Apesar de todos os dados estarem com uma forma de medição homogênea ao longo do período estudado e em uma única unidade, para a aplicação dos algoritmos de mineração os atributos devem estar classificados em faixas identificadas, devido a grande variação nos valores dos atributos abordados. Um exemplo é a variação entre o preço da ação de R$ 4,61 do dia 1/04/2003 com o preço da mesma ação no dia 01/07/2008 de R$ 45,50. Além disso com o objetivo de melhor identificar a faixa mais adequada ao contexto aplicado optou-se trabalhar com diversas faixas de classificação de diferentes tamanhos.

Para escolha da quantidade de faixas e seus respectivos tamanhos, apesar de ter sido uma decisão aleatória, pois a variação percentual diária é diversificada, alguns fatores foram considerados. Com isso, foram gerados diversos arquivos com faixas variadas. O primeiro foi a distribuição dos dados de cada atributo dentro de uma faixa de valores de -100% a 100% de variação.

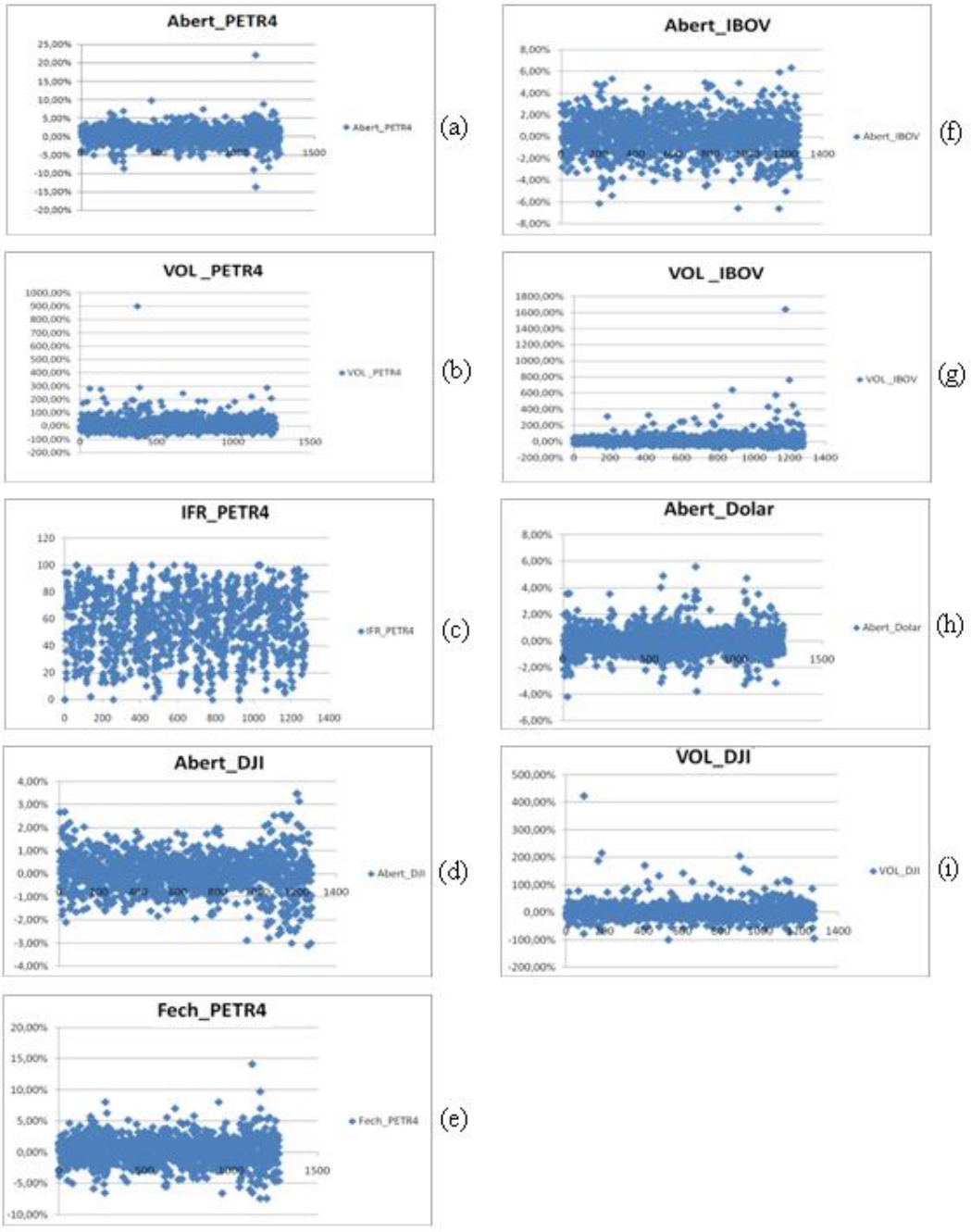


**Figura 4.7:** Gráficos de dispersão dos atributos dos ativos.
**Fonte:** Do autor.

Pelas concentrações observadas nos gráficos da figura 4.7, nota-se que a grande maioria dos dados apresentados nos atributos fechamento de PETR4 (e), abertura de Petr4 (a), abertuda de Dow Jones (d), abertura do dólar (h) e abertura de Ibovespa (f), tiveram sua concentração de dados em uma faixa dominante de 5% a -5%. Este fato fez com quem fossem definidas algumas faixas focando este escopo, que foram utilizadas nos arquivos. A primeira,

e válida para todos os atributos foi a definição dos percentuais com relação a seu valor sendo maior que zero e menor que zero.

Outra classificação, foi a separação de faixas a cada 3%, de -12% a +12% em cada um dos atributos. Para valores fora desta faixa, foram criadas faixas para valores acima de +12% ( Maior12) e para valores abaixo de -12% (Menor12%). Atributos como volumes de Petrobrás, Ibovespa, Dow Jones e IFR de Petrobrás, tiveram suas faixas definidas a cada 50% de -100% a +100%. Valores acima destas faixas estabelecidas foram definidos como Maior100 e Menor -100, respectivamente para valores acima de 100% e para valores abaixo de -100%.

Uma terceira forma de classificação foi a separação de faixas a cada 5%, de -10% a +10% em cada um dos atributos. Para valores fora desta faixa, foram criadas faixas para valores acima de +10% (Maior10) e para valores abaixo de -10% (Menor10). Atributos como volumes de Petrobrás, Ibovespa, Dow Jones e IFR de Petrobrás tiveram suas faixas mantidas em faixas a cada 50% como explicado anteriormente. As faixas com a utilização dos limites de -12 a +12% e -10% a +10%, assim como as faixas maior e menor que zero tem como objetivo a concentração dos dados em faixas onde a variação diária é maior que 1%.

Por fim, foi criada uma quarta faixa de classificação dos dados, onde foi a separação de faixas é feita a cada 1%, de -7% a +7% em cada um dos atributos. Para valores fora desta faixa, foram criadas faixas para valores acima de +7% ( Maior7) e para valores abaixo de -7% (Menor-7). Atributos de volumes de Petrobrás, Ibovespa, Dow Jones e IFR de Petrobrás, tiveram suas faixas mantidas em faixas a cada 50% como explicado anteriormente.

Optou-se por não fazer uma classificação direta por cada ponto percentual porque isto geraria muitas faixas em cada atributo analisado, e com isso, a criação de muitas ramificações na técnica de árvores de decisão, por exemplo.

### 4.1.4 Novos atributos

Algumas situações exigem a criação de novos atributos, facilitando e focando o processo de descoberta de conhecimento (REZENDE,2005). No caso do estudo das ações da Bovespa, além dos atributos puros retirados da base de dados, há a necessidade de criação de novos atributos para identificação de relações existentes entre os dados. Estes atributos podem ser obtidos através de cálculos relacionando itens da base de dados, através da variação entre os registros.

O uso de atributos como valores de abertura e fechamento dos dias que antecedem a data de registros é utilizada de forma gráfica na análise técnica. Como no exemplo da figura 4.8, onde o gráfico mensal traz amostras em períodos diários, onde o usuário interpreta os dados diários e busca padrões lógicos ou repetitivos na figura do gráfico. Esta comparação é a inspiração para o uso de valores de abertura ou fechamento de uma seqüência de dias. Na figura 4.8 , se a abertura foi menor que o fechamento do dia, o registro aparece em vermelho, indicando queda da cotação da ação ou, no caso do preço de fechamento ser superior ao preço de abertura, em verde, indicando elevação na cotação da ação.


**Figura 4.8:** Exemplo de amostra diária na análise técnica tradicional
**Fonte:** Do autor

Já que o objetivo é a identificação de padrões que traga ao usuário lucros ao fim de certos períodos, o atributo fechamento diário da ação da Petrobrás foi selecionado como o atributo a ser distribuído ao longo do tempo, pois com este atributo o usuário terá tempo hábil para efetuar suas negociações, comprando ou vendendo ações.

O período abordado neste trabalho permitiu a criação de diversos atributos a serem testados, como o estudo para o valor de fechamento do dia posterior ( D+1), o fechamento para dois ( D+2), três (D+3),cinco ( D+5), dez (D+10) e quinze ( D+15) dias posteriores à data do registro, com base nos fechamentos de dez dias úteis anteriores ao registro. Por exemplo. O registro de 10/10/2008 (sexta-feira) leva como atributos os fechamentos diários desde o dia 03/10/2008, buscando a identificação de padrões para o dia 13/10/2008 (segunda-feira, D+1), ou para 14/10/2008 (D+2), ou para 15/10/2008 (D+3), ou para 17/10/2008 ( D+5), ou para 24/10/2008 (D+10), ou para 31/10/2008 ( D+15).

Esta forma de estudo visa eliminar a maior dificuldade para uma analista técnico, que é a análise das tendências no futuro, ou seja, na ponta direita do gráfico, como mostrado na figura 4.8. O investidor está situado, na linha de tempo do gráfico sempre na amostra mais recente, sendo que o futuro é desconhecido, e esta tendência, o rumo do preço da ação, é o que trará sucesso ou não ao investidor. Aplicar *data mining* para eliminar este problema é uma nova forma de tratar este assunto, pois nas ferramentas gráficas hoje existentes esta tendência é estimada com o uso de estudos estatísticos, como médias móveis, por exemplo.

Buscou-se neste trabalho a definição da faixa de variação do valor de fechamento da ação entre uma data com a data abordada, com base nos períodos de tempo citados anteriormente ( dia posterior ( D+1), dois dias ( D+2), três (D+3), cinco ( D+5), dez (D+10) e quinze ( D+15) dias posteriores). Além da variação do preço de fechamento, buscou-se identificar a tendência positiva ou negativa entre o preço de fechamento do dia abordado e o fechamento de outro dia dos períodos citados acima.

O objetivo desta abordagem é identificar padrões que mostrem ao investidor pessoa física , que apesar das variações positivas ou negativas a curto prazo, e dentro do período estudado ( no caso, D+1, ou D+3, ou D+10, por exemplo), com relação a data abordada há

uma probalilidade determinada do preço da ação estar acima ou abaixo do valor atual, deixando o investidor mais tranquilo com relação à volatividade do mercado.

Como exemplo, no gráfico da figura 4.9, poderia ser estimada a variação entre o dia 8 de setembro e o dia 22 de setembro, mostrando uma variação positiva. Com isso, o investidor passaria tranquilo por todo período de queda entre estas duas datas, pois já teria uma probalilidade de alta do mercado estimada em valores históricos.


**Figura 4.9:** Exemplo gráfico de variação diária da Petrobrás.
**Fonte:** Do autor.

Desta forma, foram criados arquivos de base para estudos de tendências e variações. Como estes dados são relacionados a registros anteriores e posteriores ao estudado, alguns registros de início e fim da base formada ficam sem alguns valores nos atributos, como mostrado na figura 4.10, o início de um arquivo. Logo, a opção para tratamento desta não conformidade foi a exclusão dos registros onde faltavam dados. Nas figuras 4.10 e 4.11, são mostradas as bases de variação e tendências formadas.

| Variação D-5 | Variação D-4 | Variação D-3 | Variação D-2 | Abertura | Fechamento | Fechamento D+1 | Fechamento D+2 | Fechamento D+3 | Fechamento D+5 | Fechamento D+10 | Fechamento D+15 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | #DIV/0! | #DIV/0! | #DIV/0! | 2,82% | 1,49% | 1,26% | 1,67% | -1,26% | -4,39% | -4,18% | -1,88% |
| #DIV/0! | #DIV/0! | #DIV/0! | 2,76% | 1,27% | 1,26% | 0,41% | -2,48% | -6,20% | -5,58% | -2,89% | -2,27% |
| #DIV/0! | #DIV/0! | 3,18% | 1,67% | 1,04% | 0,41% | -2,88% | -6,58% | -5,97% | -5,76% | -2,06% | -0,41% |
| #DIV/0! | 0,21% | -1,26% | -2,48% | 1,86% | -2,88% | -3,81% | -3,18% | -3,18% | -2,97% | 1,91% | 4,45% |
| -3,61% | -5,02% | -6,20% | -6,58% | -4,05% | -3,81% | 0,66% | 0,66% | 0,88% | 1,76% | 4,41% | 10,57% |
| -4,39% | -5,58% | -5,97% | -3,18% | -3,59% | 0,66% | 0,00% | 0,22% | 0,22% | 0,22% | 2,63% | 12,04% |
| -5,58% | -5,97% | -3,18% | 0,66% | 0,22% | 0,00% | 0,22% | 0,22% | 1,09% | 2,84% | 3,50% | 11,38% |
| -5,76% | -2,97% | 0,88% | 0,22% | 0,00% | 0,22% | 0,00% | 0,87% | 0,00% | 3,93% | 5,68% | 12,66% |
| -2,97% | 0,88% | 0,22% | 0,22% | 0,44% | 0,00% | 0,87% | 0,00% | 2,62% | 5,02% | 7,64% | 10,48% |
| 1,76% | 1,09% | 1,09% | 0,87% | 0,00% | 0,87% | -0,87% | 1,73% | 3,03% | 2,60% | 8,66% | 12,55% |
| 0,22% | 0,22% | 0,00% | 0,00% | 0,87% | -0,87% | 2,62% | 3,93% | 5,02% | 2,40% | 11,79% | 15,72% |
| 2,84% | 2,62% | 2,62% | 1,73% | 0,22% | 2,62% | 1,28% | 2,34% | 0,85% | 0,64% | 8,30% | 14,47% |
| 3,93% | 3,93% | 3,03% | 3,93% | 0,00% | 1,28% | 1,05% | -0,42% | -1,47% | 1,68% | 8,40% | 12,18% |
| 5,02% | 4,11% | 5,02% | 2,34% | 1,94% | 1,05% | -1,46% | -2,49% | -1,66% | 2,49% | 5,20% | 12,06% |
| 2,60% | 3,49% | 0,85% | -0,42% | 1,27% | -1,46% | -1,05% | -0,21% | 2,11% | 5,91% | 9,70% | 13,08% |
| 2,40% | -0,21% | -1,47% | -2,49% | -1,25% | -1,05% | 0,85% | 3,20% | 5,12% | 9,17% | 13,01% | 10,87% |
| 0,64% | -0,63% | -1,66% | -0,21% | -0,84% | 0,85% | 2,33% | 4,23% | 6,13% | 7,61% | 13,74% | 9,94% |
| 1,68% | 0,62% | 2,11% | 3,20% | 2,13% | 2,33% | 1,86% | 3,72% | 5,79% | 6,61% | 10,33% | 9,30% |
| 2,49% | 4,01% | 5,12% | 4,23% | 2,50% | 1,86% | 1,83% | 3,85% | 3,25% | 2,64% | 9,33% | 7,91% |
| 5,91% | 7,04% | 6,13% | 3,72% | 0,20% | 1,83% | 1,99% | 1,39% | 2,79% | 3,59% | 6,77% | 4,38% |
| 9,17% | 8,25% | 5,79% | 3,85% | 1,62% | 1,99% | -0,59% | 0,78% | -1,17% | 3,52% | 1,56% | 2,93% |
| 7,61% | 5,17% | 3,25% | 1,39% | 1,40% | -0,59% | 1,38% | -0,59% | 2,16% | 5,70% | 2,16% | 4,13% |
| 6,61% | 4,67% | 2,79% | 0,78% | 0,00% | 1,38% | -1,94% | 0,78% | 2,71% | 3,49% | 2,52% | 1,16% |
| 2,64% | 0,80% | -1,17% | -0,59% | 1,18% | -1,94% | 2,77% | 4,74% | 6,32% | 6,52% | 5,14% | 3,16% |
| 3,59% | 1,56% | 2,16% | 0,78% | -1,17% | 2,77% | 1,92% | 3,46% | 2,69% | 3,08% | 0,77% | -0,19% |
| 3,52% | 4,13% | 2,71% | 4,74% | 2,36% | 1,92% | 1,51% | 0,75% | 1,70% | -1,89% | -0,57% | -2,26% |
| 5,70% | 4,26% | 6,32% | 3,46% | 1,92% | 1,51% | -0,74% | 0,19% | -0,37% | -3,35% | -1,49% | -0,74% |
| 3,49% | 5,53% | 2,69% | 0,75% | 2,64% | -0,74% | 0,94% | 0,37% | -2,62% | -0,94% | -2,25% | 0,00% |
| 6,52% | 3,65% | 1,70% | 0,19% | -1,84% | 0,94% | -0,56% | -3,53% | -3,53% | -1,30% | -3,15% | -0,37% |
| 3,08% | 1,13% | -0,37% | 0,37% | 3,56% | -0,56% | -2,99% | -2,99% | -1,31% | -2,24% | -3,17% | -0,56% |
| -1,89% | -3,35% | -2,62% | -3,53% | -4,16% | -2,99% | 0,00% | 1,73% | 2,31% | 1,35% | -0,38% | 0,19% |
| -3,35% | -2,62% | -3,53% | -2,99% | -1,51% | 0,00% | 1,73% | 2,31% | 0,77% | 1,92% | 2,69% | 1,92% |
| -0,94% | -1,86% | -1,31% | 1,73% | 0,00% | 1,73% | 0,57% | -0,95% | -0,38% | -1,32% | 0,95% | 1,70% |

**Figura 4.10** Exemplo de arquivo base para variações.
**Fonte:** Do autor.

| FechaD-5 | FechaD-4 | FechD-3 | FechD-2 | FechD-1 | Fech | Abert | FechD+1 | FechaD+2 | FechD+3 | FechD+5 | FechD+10 | FechaD+15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | #REF! | #REF! | #REF! | 0,00% | 1,49% | 2,82% | 1,26% | 0,41% | -2,88% | 0,66% | -0,87% | -1,05% |
| #REF! | #REF! | #REF! | 0,00% | 1,49% | 1,26% | 1,27% | 0,41% | -2,88% | -3,81% | 0,00% | 2,62% | 0,85% |
| #REF! | #REF! | 0,00% | 1,49% | 1,26% | 0,41% | 1,04% | -2,88% | -3,81% | 0,66% | 0,22% | 1,28% | 2,33% |
| #REF! | 0,00% | 1,49% | 1,26% | 0,41% | -2,88% | 1,86% | -3,81% | 0,66% | 0,00% | 0,00% | 1,05% | 1,86% |
| 0,00% | 1,49% | 1,26% | 0,41% | -2,88% | -3,81% | -4,05% | 0,66% | 0,00% | 0,22% | 0,87% | -1,46% | 1,83% |
| 1,49% | 1,26% | 0,41% | -2,88% | -3,81% | 0,66% | -3,59% | 0,00% | 0,22% | 0,00% | -0,87% | -1,05% | 1,99% |
| 1,26% | 0,41% | -2,88% | -3,81% | 0,66% | 0,00% | 0,22% | 0,22% | 0,00% | 0,87% | 2,62% | 0,85% | -0,59% |
| 0,41% | -2,88% | -3,81% | 0,66% | 0,00% | 0,22% | 0,00% | 0,00% | 0,87% | -0,87% | 1,28% | 2,33% | 1,38% |
| -2,88% | -3,81% | 0,66% | 0,00% | 0,22% | 0,00% | 0,44% | 0,87% | -0,87% | 2,62% | 1,05% | 1,86% | -1,94% |
| -3,81% | 0,66% | 0,00% | 0,22% | 0,00% | 0,87% | 0,00% | -0,87% | 2,62% | 1,28% | -1,46% | 1,83% | 2,77% |
| 0,66% | 0,00% | 0,22% | 0,00% | 0,87% | -0,87% | 0,87% | 2,62% | 1,28% | 1,05% | -1,05% | 1,99% | 1,92% |
| 0,00% | 0,22% | 0,00% | 0,87% | -0,87% | 2,62% | 0,22% | 1,28% | 1,05% | -1,46% | 0,85% | -0,59% | 1,51% |
| 0,22% | 0,00% | 0,87% | -0,87% | 2,62% | 1,28% | 0,00% | 1,05% | -1,46% | -1,05% | 2,33% | 1,38% | -0,74% |
| 0,00% | 0,87% | -0,87% | 2,62% | 1,28% | 1,05% | 1,94% | -1,46% | -1,05% | 0,85% | 1,86% | -1,94% | 0,94% |
| 0,87% | -0,87% | 2,62% | 1,28% | 1,05% | -1,46% | 1,27% | -1,05% | 0,85% | 2,33% | 1,83% | 2,77% | -0,56% |
| -0,87% | 2,62% | 1,28% | 1,05% | -1,46% | -1,05% | -1,25% | 0,85% | 2,33% | 1,86% | 1,99% | 1,92% | -2,99% |
| 2,62% | 1,28% | 1,05% | -1,46% | -1,05% | 0,85% | -0,84% | 2,33% | 1,86% | 1,83% | -0,59% | 1,51% | 0,00% |
| 1,28% | 1,05% | -1,46% | -1,05% | 0,85% | 2,33% | 2,13% | 1,86% | 1,83% | 1,99% | 1,38% | -0,74% | 1,73% |
| 1,05% | -1,46% | -1,05% | 0,85% | 2,33% | 1,86% | 2,50% | 1,83% | 1,99% | -0,59% | -1,94% | 0,94% | 0,57% |
| -1,46% | -1,05% | 0,85% | 2,33% | 1,86% | 1,83% | 0,20% | 1,99% | -0,59% | 1,38% | 2,77% | -0,56% | -1,50% |
| -1,05% | 0,85% | 2,33% | 1,86% | 1,83% | 1,99% | 1,62% | -0,59% | 1,38% | -1,94% | 1,92% | -2,99% | 0,57% |
| 0,85% | 2,33% | 1,86% | 1,83% | 1,99% | -0,59% | 1,40% | 1,38% | -1,94% | 2,77% | 1,51% | 0,00% | 0,57% |
| 2,33% | 1,86% | 1,83% | 1,99% | -0,59% | 1,38% | 0,00% | -1,94% | 2,77% | 1,92% | -0,74% | 1,73% | -1,51% |
| 1,86% | 1,83% | 1,99% | -0,59% | 1,38% | -1,94% | 1,18% | 2,77% | 1,92% | 1,51% | 0,94% | 0,57% | 0,00% |
| 1,83% | 1,99% | -0,59% | 1,38% | -1,94% | 2,77% | -1,17% | 1,92% | 1,51% | -0,74% | -0,56% | -1,50% | -0,57% |
| 1,99% | -0,59% | 1,38% | -1,94% | 2,77% | 1,92% | 2,36% | 1,51% | -0,74% | 0,94% | -2,99% | 0,57% | -0,19% |
| -0,59% | 1,38% | -1,94% | 2,77% | 1,92% | 1,51% | 1,92% | -0,74% | 0,94% | -0,56% | 0,00% | 0,57% | 3,09% |
| 1,38% | -1,94% | 2,77% | 1,92% | 1,51% | -0,74% | 2,64% | 0,94% | -0,56% | -2,99% | 1,73% | -1,51% | 0,00% |
| -1,94% | 2,77% | 1,92% | 1,51% | -0,74% | 0,94% | -1,84% | -0,56% | -2,99% | 0,00% | 0,57% | 0,00% | 0,56% |
| 2,77% | 1,92% | 1,51% | -0,74% | 0,94% | -0,56% | 3,56% | -2,99% | 0,00% | 1,73% | -1,50% | -0,57% | -0,74% |
| 1,92% | 1,51% | -0,74% | 0,94% | -0,56% | -2,99% | -4,16% | 0,00% | 1,73% | 0,57% | 0,57% | -0,19% | -2,25% |
| 1,51% | -0,74% | 0,94% | -0,56% | -2,99% | 0,00% | -1,51% | 1,73% | 0,57% | -1,50% | 0,57% | 3,09% | 1,73% |
| -0,74% | 0,94% | -0,56% | -2,99% | 0,00% | 1,73% | 0,00% | 0,57% | -1,50% | 0,57% | -1,51% | 0,00% | 1,51% |

**Figura 4.11:** Exemplo de arquivo base de tendência.
**Fonte:** Do autor.

Destas tabelas foram extraídos atributos e classificados de acordo com a faixa estipulada, que juntamente com os outros atributos originais formaram os arquivos base para a geração dos arquivos arff. Este tipo de arquivo contém os dados já tratados para a aplicação do algoritmo de mineração.

| Abert_Petr4 | VOL _Petr4 | IFR Petr4 | Abert_Ibov | VOL_Ibov | Abert_Dol | Abert_ Djonis | VOL _ Djonis | FechaD-5 | FechaD-4 | FechD-3 | FechD-2 | FechD-1 | Fech | FechD+1 | FechaD+2 | FechD+3 | FechD+5 | FechD+10 | FechaD+1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -3,59% | -33,08% | 49,2063 | -2,81% | -14,56% | 0,38% | 0,00% | -21,26% | 0,014862 | 0,012552 | 0,00413 | -2,88% | -3,81% | 0,66% | 0,00% | 0,22% | 0,00% | -0,87% | -1,05% | 1,99% |
| 0,22% | 39,43% | 36 | 0,00% | -8,47% | 1,99% | -1,21% | 4,08% | 0,012552 | 0,004132 | -0,02881 | -3,81% | 0,66% | 0,00% | 0,22% | 0,00% | 0,87% | 2,62% | 0,85% | -0,59% |
| 0,00% | -40,92% | 27,2727 | -1,44% | -0,88% | -0,22% | 0,30% | -6,48% | 0,004132 | -0,02881 | -0,03814 | 0,66% | 0,00% | 0,22% | 0,00% | 0,87% | -0,87% | 1,28% | 2,33% | 1,38% |
| 0,44% | 4,82% | 15,7895 | 0,96% | 1,52% | -0,93% | -0,23% | 25,37% | -0,02881 | -0,03814 | 0,00661 | 0,00% | 0,22% | 0,00% | 0,87% | -0,87% | 2,62% | 1,05% | 1,86% | -1,94% |
| 0,00% | -22,08% | 20 | 1,19% | 39,94% | -1,10% | 1,75% | 21,76% | -0,03814 | 0,006608 | 0 | 0,22% | 0,00% | 0,87% | -0,87% | 2,62% | 1,28% | -1,46% | 1,83% | 2,77% |
| 0,87% | 89,08% | 26,6667 | 2,14% | -13,58% | -2,63% | 0,70% | -19,72% | 0,006608 | 0 | 0,00219 | 0,00% | 0,87% | -0,87% | 2,62% | 1,28% | 1,05% | -1,05% | 1,99% | 1,92% |
| 0,22% | -32,39% | 83,3333 | -0,56% | -21,56% | -0,36% | -1,78% | -28,20% | 0 | 0,002188 | 0 | 0,87% | -0,87% | 2,62% | 1,28% | 1,05% | -1,46% | 0,85% | -0,59% | 1,51% |
| 0,00% | 47,79% | 85,1852 | 2,83% | 9,48% | 0,52% | 0,86% | 42,27% | 0,002188 | 0 | 0,00873 | -0,87% | 2,62% | 1,28% | 1,05% | -1,46% | -1,05% | 2,33% | 1,38% | -0,74% |
| 1,94% | -3,24% | 87,5 | 0,51% | -29,14% | -0,65% | 1,90% | -44,91% | 0 | 0,008734 | -0,00866 | 2,62% | 1,28% | 1,05% | -1,46% | -1,05% | 0,85% | 1,86% | -1,94% | 0,94% |
| 1,27% | -23,38% | 71,0526 | -0,68% | 19,14% | -1,57% | 0,32% | 73,54% | 0,008734 | -0,00866 | 0,0262 | 1,28% | 1,05% | -1,46% | -1,05% | 0,85% | 2,33% | 1,83% | 2,77% | -0,56% |
| -1,25% | -8,89% | 62,7907 | -1,99% | -13,67% | 0,96% | -0,85% | -5,94% | -0,00866 | 0,026201 | 0,01277 | 1,05% | -1,46% | -1,05% | 0,85% | 2,33% | 1,86% | 1,99% | 1,92% | -2,99% |
| -0,84% | -34,51% | 62,7907 | 0,03% | 28,38% | -1,15% | -1,58% | -5,28% | 0,026201 | 0,012766 | 0,0105 | -1,46% | -1,05% | 0,85% | 2,33% | 1,86% | 1,83% | -0,59% | 1,51% | 0,00% |
| 2,13% | 169,78% | 76 | 2,82% | 46,07% | -2,17% | 2,00% | 11,25% | 0,012766 | 0,010504 | -0,01455 | -1,05% | 0,85% | 2,33% | 1,86% | 1,83% | 1,99% | 1,38% | -0,74% | 1,73% |
| 2,50% | -9,82% | 74,4681 | 1,95% | -27,27% | -2,08% | 0,34% | 3,79% | 0,010504 | -0,01455 | -0,01055 | 0,85% | 2,33% | 1,86% | 1,83% | 1,99% | -0,59% | -1,94% | 0,94% | 0,57% |
| 0,20% | -37,30% | 76 | -1,17% | -23,79% | 2,12% | -0,56% | -11,72% | -0,01455 | -0,01055 | 0,00853 | 2,33% | 1,86% | 1,83% | 1,99% | -0,59% | 1,38% | 2,77% | -0,56% | -1,50% |
| 1,62% | 15,22% | 78,1818 | 2,11% | 34,83% | 0,72% | 1,53% | -8,59% | -0,01055 | 0,008529 | 0,02326 | 1,86% | 1,83% | 1,99% | -0,59% | 1,38% | -1,94% | 1,92% | -2,99% | 0,57% |
| 1,40% | 45,74% | 84,3137 | -0,01% | -3,09% | 3,55% | -0,60% | 14,37% | 0,008529 | 0,023256 | 0,0186 | 1,83% | 1,99% | -0,59% | 1,38% | -1,94% | 2,77% | 1,51% | 0,00% | 0,57% |
| 0,00% | 30,65% | 94,3396 | -1,24% | 26,42% | -1,96% | 0,63% | -6,42% | 0,023256 | 0,018595 | 0,01826 | 1,99% | -0,59% | 1,38% | -1,94% | 2,77% | 1,92% | -0,74% | 1,73% | -1,51% |
| 1,18% | -26,97% | 77,9661 | 2,36% | -18,42% | 0,00% | -0,31% | -8,53% | 0,018595 | 0,018256 | 0,01992 | -0,59% | 1,38% | -1,94% | 2,77% | 1,92% | 1,51% | 0,94% | 0,57% | 0,00% |
| -1,17% | 6,10% | 79,0323 | -0,15% | 53,13% | -4,20% | -0,77% | 3,43% | 0,018256 | 0,01992 | -0,00586 | 1,38% | -1,94% | 2,77% | 1,92% | 1,51% | -0,74% | -0,56% | -1,50% | -0,57% |
| 2,36% | -23,80% | 79,3651 | 1,98% | -39,16% | -0,24% | 2,68% | 5,53% | 0,01992 | -0,00586 | 0,01375 | -1,94% | 2,77% | 1,92% | 1,51% | -0,74% | 0,94% | -2,99% | 0,57% | -0,19% |
| 1,92% | 44,74% | 79,0323 | 0,88% | 38,18% | -0,45% | 0,03% | -7,84% | -0,00586 | 0,013752 | -0,01938 | 2,77% | 1,92% | 1,51% | -0,74% | 0,94% | -0,56% | 0,00% | 0,57% | 3,09% |
| 2,64% | -25,78% | 69,6429 | 0,80% | -19,27% | 1,93% | -0,57% | 4,20% | 0,013752 | -0,01938 | 0,02767 | 1,92% | 1,51% | -0,74% | 0,94% | -0,56% | -2,99% | 1,73% | -1,51% | 0,00% |
| -1,84% | 14,39% | 75,8621 | 0,27% | -0,84% | 1,17% | -0,28% | 5,53% | -0,01938 | 0,027668 | 0,01923 | 1,51% | -0,74% | 0,94% | -0,56% | -2,99% | 0,00% | 0,57% | 0,00% | 0,56% |
| 3,56% | 64,09% | 68,5185 | -2,38% | -1,97% | 0,51% | 0,71% | 14,61% | 0,027668 | 0,019231 | 0,01509 | -0,74% | 0,94% | -0,56% | -2,99% | 0,00% | 1,73% | -1,50% | -0,57% | -0,74% |
| -4,16% | -37,64% | 61,6667 | 0,27% | 5,39% | -0,68% | -0,39% | -9,72% | 0,019231 | 0,015094 | -0,00743 | 0,94% | -0,56% | -2,99% | 0,00% | 1,73% | 0,57% | 0,57% | -0,19% | -2,25% |
| -1,51% | -8,51% | 50 | -3,23% | -12,54% | 1,63% | -2,10% | 18,86% | 0,015094 | -0,00743 | 0,00936 | -0,56% | -2,99% | 0,00% | 1,73% | 0,57% | -1,50% | 0,57% | 3,09% | 1,73% |
| 0,00% | -32,03% | 48,8889 | -0,02% | -2,23% | 1,20% | -0,10% | 20,58% | -0,00743 | 0,009363 | -0,00557 | -2,99% | 0,00% | 1,73% | 0,57% | -1,50% | 0,57% | -1,51% | 0,00% | 1,51% |

**Figura 4.12:** Exemplo de arquivo CSV para conversão ARFF
**Fonte:** Do autor.

Como software de mineração, é utilizado o Weka (Waikato Environment for Knowledge Analysis), de código aberto, portável e que possui algoritmos das duas técnicas utilizadas neste trabalho. Esta ferramenta de mineração já foi utilizada por Gonchoroski, (2007) na tarefa de classificação. Detalhes e informações sobre esta ferramenta, inclusive o download pode ser realizado a partir de http://www.cs.waikato.ac.nz/ml/weka/.

Neste ponto todos os dados estão com a mesma classificação e formam a base geral utilizada. Para as tarefas de classificação o Weka disponibiliza em seu modo gráfico, quatro formas de testes, sendo elas: a ativação da validação cruzada, o percentual de separação, a utilização de registros do próprio arquivo de treinamento como teste e por último a utilização de um arquivo de teste, criado separadamente. Este arquivo de teste é um arquivo que deve conter as mesmas características do arquivo de treinamento. Deve possuir um nome, a identificação dos atributos e a parte de dados, que serão utilizados para testes. É importante ressaltar que o arquivo de teste deve possuir o mesmo número de atributos do arquivo de treinamento, caso contrário o programa não permite a utilização dos dois arquivos.

Para evitar que o teste tanto do algoritmo de árvores de decisão, como do algoritmo de redes neurais ficasse “viciado”, optou-se por utilizar um arquivo de testes separado, ou seja, diferente do arquivo de treinamento. A geração dos arquivos de testes tomava como base a planilha de atributos em percentual com cerca de 1270 registros que através de uma função randômica criada, selecionava linhas de forma aleatória e movia para outra planilha. Esta rotina separava 70% dos registros para a base de treinamento e o restante, 30% foram utilizados para o arquivo de teste, e salvos novamente com o formato csv.

O trabalho está sendo moldado dentro do Excel, devido ao formato dos dados exportados já estarem em um formato csv, tanto para a criação dos arquivos com extensão arff como para a normalização dos dados de todos os atributos e para a criação das faixas, foram montadas macros em VBA, sendo esta a linguagem de programação utilizada dentro do pacote Office.

Atendendo ao número de faixas criadas para cada atributo foi estabelecida uma função em VBA que através de uma rotina analisava cada valor da coluna selecionada e enquadrava este valor dentro de uma faixa nominal estabelecida. Esta função estava vinculada a um formulário, sendo este a *interface* com o usuário.

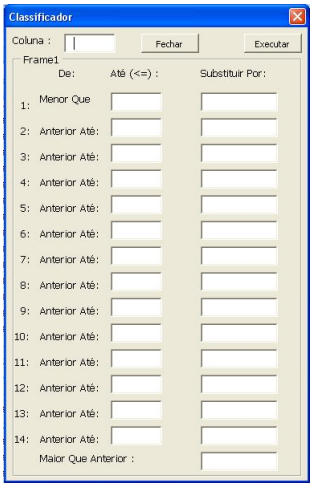


**Figura 4.13:** Imagem do Classificador criado
**Fonte:**Do autor

Através deste classificador, os arquivos de teste e treinamento foram colocados em faixas, e após esta tarefa, o arquivo com extensão csv gerado está apto a ser convertido para um formato texto e desta forma ser utilizado como a parte de dados de um arquivo arff.

### 4.1.5 Formação dos arquivos ARFF

A geração dos arquivos de entrada para o algoritmo de mineração deve ser realizada identificando antes dos dados, os atributos que estão presentes no arquivo e as faixas de valores válidas, além do nome do arquivo.

De forma manual os atributos tiveram suas faixas listadas e colocadas nos arquivos de teste e treinamento correspondentes, acompanhados da tag *@attribute*, que é o identificador de atributos do weka. Também foram colocados o nome do arquivo com a tag *@relation*, que é a identificação de nomes de arquivos do weka. Na figura 4.14 a seguir, um exemplo de como ficaram os padrões dos arquivos arff. Neste exemplo, as faixas para os atributos exceto volume de Petrobrás, IFR da Petrobrás volume da Bovespa e volume de Dow Jones tiveram apenas dois valores. Maior0, ou seja, positivo e Menor0, negativo.

```
@relation treMaiorOuMenor0

@attribute Abert_Petr4 {Menor0,Maior0}
@attribute VOL_Petr4 {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute IFR_Petr4 {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute Abert_IBOV {Menor0,Maior0}
@attribute VOL_IBOV {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute Abert_DOLAR {Menor0,Maior0}
@attribute Abert_DJI {Menor0,Maior0}
@attribute VOL_DJI {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute Fechamento {Menor0,Maior0}
@attribute FechamentoPost {Menor0,Maior0}
@attribute VarD+2 {Menor0,Maior0}

@data

Maior0,-50a0,0a50,Maior0,-50a0,Menor0,Maior0,-50a0,Maior0,Maior0,Menor0
Maior0,0a50,50a100,Maior0,50a100,Menor0,Maior0,-50a0,Menor0,Menor0,Menor0
Menor0,-50a0,0a50,Menor0,0a50,Menor0,Maior0,0a50,Menor0,Maior0,Maior0
Maior0,0a50,0a50,Maior0,-50a0,Maior0,Menor0,0a50,Maior0,Maior0,Maior0
Maior0,0a50,0a50,Maior0,0a50,Menor0,Menor0,0a50,Maior0,Maior0,Maior0
Maior0,-50a0,0a50,Maior0,0a50,Menor0,Maior0,0a50,Maior0,Menor0,Maior0
Maior0,50a100,0a50,Maior0,-50a0,Menor0,Maior0,-50a0,Menor0,Maior0,Maior0
```

**Figura 4.14:** Exemplo de cabeçalho de arquivo arff.
**Fonte:** Do autor.

O estudo das faixas de valores de cada atributo e as possíveis classes em que estes poderiam ser enquadrados, a limpeza e discretização fazem parte da etapa de pré-processamento dos dados de base, sendo esta uma etapa anterior a mineração de dados, ou seja, a aplicação do algoritmo. Com base nos dados analisados foram então criados diversos arquivos de treinamento e testes, tornando os dados estudados adequados à aplicação do(s) algoritmo(s) de mineração.

## 4.2 Os Arquivos Criados

Os arquivos criados tinham objetivos diferentes. Os arquivos “Treinamento Tendência D+1”,” Treinamento Tendência D+2”, “ Treinamento Tendência D+3”, “ Treinamento Tendência D+5”, “ Treinamento Tendência D+10” e “ Treinamento Tendência D+15”, foram criados para identificação de relações entre os fechamentos diários passados e futuros em conjunto com os valores do dia de abertura, volume, IFR e fechamento da Petrobrás além do uso de valores como índice e volume de negociações da Bovespa, Dow Jones e cotação do dólar comercial. O período objetivado é, respectivamente, de um, dois, três, cinco, dez e quinze dias posteriores ao dia atual de cada registro. Na tabela a seguir o exemplo de um registro classificado objetivando a relação com 3 dias posteriores ao dia atual. O dia atual tem como atributos o valor de fechamento e abertura do dia. Estes atributos são comparados com os fechamentos de cinco dias anteriores, relacionados com os dados de três dias posteriores.

**Tabela 4.1:** Registro classificado objetivando a relação com 3 dias posteriores ao dia atual.

| FechaD-5 | FechaD-4 | FechD-3 | FechD-2 | FechD-1 | Fech | Abert | FechD+1 | FechaD+2 | FechD+3 |
|---|---|---|---|---|---|---|---|---|---|
| 1a2 | 1a2 | 0a1 | -3a-2 | Menor-3 | 0a1 | Menor-3 | 0a1 | 0a1 | 0a1 |

**Fonte:** Do autor.

Da mesma forma, foram criados arquivos para estudo de padrões entre as variações diárias com base em dados do passado e do futuro, nos mesmos períodos citados anteriormente, tendo como referência os valores do dia atual do registro, inclusive com uso dos mesmos atributos de valores, preços e índices diários. Estes arquivos, receberam o nome de “Treinamento Variação D+1”, “Treinamento Variação D+2”, “Treinamento Variação D+3”, “Treinamento Variação D+5”, “Treinamento Variação D+10”, “Treinamento Variação D+15”.

Os arquivos treinamentoHoje, treHoje+5, treinamentoHoje+10, treHoje+15, treHoje+20, visavam obter padrões com base nos fechamentos dos últimos 10 dias e na relação entre estes dias com dia posterior ou 5,10,15 e 20 dias para frente, com faixas a cada 2% de -12 a +12, exceto volumes e IFR.

Arquivos como treinamento0a5DiaPost por exemplo, buscam a relação entre os dados do dia com os dados do dia posterior, dois e cinco dias após a data do registro, com faixas de 5% nos atributos abrangendo variações de – 10% a 10%, exceto atributos que consideram volumes e IFR onde as faixas estavam separadas em percentuais de 50%, avaliando valores de -100% a 100%. Já arquivos como treinamento0a3, treinamento0a3DiaPost, buscam a relação entre os dados do dia com os dados do dia posterior, dois e cinco dias posteriores, com faixas de 3% nos atributos abrangendo variações de – 12% a 12%, exceto para os atributos que consideram volumes e IFR onde as faixas estavam separadas em percentuais de 50%, avaliando valores de -100% a 100%.

O arquivo Treinamento, busca identificar a relação entre estes dados do dia com o dia posterior, utilizando faixas a cada 2% nos atributos abrangendo variações de – 10% a 10%, exceto atributos que consideram volumes, onde as faixas estavam separadas em percentuais de 25%, avaliando valores de -100% a 100%. Para o atributo IFR foi utilizada a faixa a cada 10%, abrangendo valores de -100% a 100%.

Treinamento_dias_da_Semana (Treinamento_Quinta, Treinamento_Sexta, Treinamento_Quarta, Treinamento_Terça, Treinamento_Segunda, Treinamento_SegundaSemD+5) são um conjunto de arquivos criados para identificar relações dos dados diários com relação ao dia posterior, dois e cinco dias à frente da data atual, considerando os dados diários conforme o dia da semana. O objetivo é observar relações que se destaquem em algum dos dias da semana, ou sazonalidades entre estes atributos.

Arquivos treVarD+5 e TreVar D+2 verificaram a relação entre os atributos diários e o fechamento posterior, dois e cinco dias posteriores, com faixas de volumes e IFR a cada 50% e 1% para os demais atributos.

Além destes foram criados arquivos com faixas de classificações amplas, como TreMaiorOuMenor0, visando identificar a relação entre os atributos diários e variação para dois e cinco dias posteriores, com atributos de volumes e IFR em faixas de 50% e demais com faixas considerando a variação do registro de forma positiva ou negativa.

Demais arquivos como treinamento5, treinamento_FaixasReclass, treinamento_FaixasReclassDiaPosterior, treinamento_PETR treinamento_SemIFR,

treinamento_SemIfrDiario e treinamento_VolReclass buscavam a relação entre o fechamento de alguns dias à frente da data atual com os dados do passado, utilizando alterações como exclusões de alguns atributos, como IFR e volume, reclassificação de faixas, etc.

Deste "*brainstorming*" de arquivos e faixas, nota-se que os arquivos TREMaiorOuMenor0D+5 e TREMaiorOuMenor0D+2 destacam-se no percentual de classificações corretas no modo de teste de *Supplied test Set*, com resultados acima de 70%. Com estas informações, e objetivando delimitar o escopo do trabalho, estes dois arquivos foram definidos como base para a aplicação e estudo da segunda técnica de mineração de dados, as redes neurais. No capítulo seguinte são descritos os testes e resultados obtidos nas etapas de teste e validação dos resultados.

# 5 TREINAMENTO E TESTES

Diversos arquivos foram criados onde as variações percentuais diárias dos atributos avaliados foram separadas em faixas de diferentes valores. Além disso, estes arquivos foram distribuídos buscando a relação entre diversos atributos alvos, como a variação do valor de fechamento da ação da PETR4 para dois, cinco ou dez dias posteriores a data atual. Desta forma objetivou-se identificar relações de destaque em cenários de curto, médio ou longo prazo. No anexo D consta a descrição de cada arquivo criado.

O trabalho propõe a utilização de duas técnicas de classificação para a definição de padrões dentro do mercado financeiro. Uma destas técnicas, as RNAs, sendo uma tecnologia considerada “caixa preta”, ou seja, onde o usuário não consegue visualizar o processamento do algoritmo, inserindo os dados de entrada e coletando os dados de saída. E outra técnica, as chamadas árvores de decisão, cujo processamento do algoritmo pode ser visualizado pelo usuário.

Na literatura de mineração de dados, autores como Ludwig e Montgomery (2007) citam que as RNA possuem a desvantagem de um longo tempo de treinamento da rede e, com base nisto, foi definida a metodologia de pesquisa. Primeiramente os arquivos com diferentes faixas nos atributos avaliados foram submetidos à classificação com a tecnologia de árvores de decisão. Os parâmetros ajustáveis desta técnica foram alterados, visando identificar os melhores valores para cada um deles. Após a identificação dos melhores parâmetros na tecnologia de árvores de decisão, os arquivos que se destacaram nesta técnica foram novamente testados com o uso de redes neurais, onde os parâmetros ajustáveis desta técnica

também foram modificados, visando identificar a RNA mais adequada. Com isso torna-se possível a comparação dos resultados das duas técnicas.

### 5.1 Treinamento e testes com árvores de decisão

Na técnica de árvores de decisão, existem alguns algoritmos disponibilizados pelo software Weka, sendo os principais os algoritmos ID3 e J48, já citados anteriormente. Para identificar as diferenças entre estes dois algoritmos, como mostrado na tabela 5.0 os arquivos criados foram testados em cada uma das técnicas, com o mesmo modo de treinamento, o *Supplied test set*. Desta forma foi possível definir que o algoritmo J48 é mais aplicável ao contexto devido a média de registros classificados de forma correta ser de cerca de 32%, superior ao resultado de 22% obtido pelo algoritmo ID3.

**Tabela 5.0:** Resultados da comparação entre os algoritmos J48 e ID3.

| Arquivo | J48 supplied test | | ID3 supplied test | | |
|---|---|---|---|---|---|
| | Correto | Erros | Correto | Incorreto | Não class |
| TREMaiorOuMenor0D+5.arff | 72% | 28% | 60% | 37% | 3% |
| TREMaiorOuMenor0D+2.arff | 76% | 24% | 61% | 36% | 3% |
| treinamento0a3.arff | 27% | 73% | 18% | 68% | 14% |
| treinamento0a3DiaPost.arff | 49% | 51% | 38% | 53% | 8% |
| treinamento0a3DiaPost_Sem IFR.arff | 49% | 51% | 38% | 53% | 9% |
| treinamento0a5.arff | 46% | 54% | 33% | 61% | 6% |
| treinamento0a5diaPost.arff | 57% | 43% | 49% | 47% | 4% |
| Treinamento5.arff | 58% | 42% | 41% | 38% | 21% |
| Treinamento.arff | 16% | 84% | 11% | 60% | 29% |
| Treinamento_FaixasReclass.arff | 41% | 59% | 30% | 59,03% | 11% |
| Treinamento_FaixasReclassDiaPosterior.arff | 48% | 52% | 45% | 48% | 6% |
| Treinamento_PETR.arff | 18% | 82% | 19% | 78% | 2% |
| Treinamento_SemIFR.arff | 13% | 87% | 11% | 62% | 27% |
| Treinamento_SemIfrDiario.arff | 20% | 80% | 11% | 54% | 36% |
| Treinamento_VolReclass.arff | 17% | 83% | 11% | 70% | 19% |
| TReVarD+2.arff | 28% | 72% | 19% | 55% | 26% |
| TReVarD+5.arff | 36% | 64% | 22% | 54% | 25% |
| TreHoje | 30% | 70% | 19% | 52% | 29% |
| TreHoje+5 | 34% | 66% | 50% | 48% | 2% |
| TreHoje+10 | 33% | 67% | 21% | 52% | 27% |
| TreHoje+15 | 32% | 68% | 18% | 52% | 30% |
| TreHoje+20 | 32% | 68% | 23% | 50% | 27% |
| Treinamento Sexta.arff | 36% | 64% | 22% | 44% | 33% |
| Treinamento_Quinta.arff | 24% | 76% | 14% | 50% | 36% |
| Treinamento_Quarta.arff | 40% | 60% | 21% | 54% | 25% |
| Treinamento_Terça.arff | 33% | 67% | 22% | 49% | 29% |
| Treinamento_Segunda.arff | 42% | 58% | 20% | 54% | 25% |
| Treinamento_SegundaSemD+5.arff | 40% | 60% | 24% | 48% | 28% |

| Arquivo | J48 supplied test | | ID3 supplied test | | |
|---|---|---|---|---|---|
| | Correto | Erros | Correto | Incorreto | Não class |
| treinamento Tendência D+1.arff | 18% | 82% | 9% | 50% | 41% |
| treinamento Tendência D+2.arff | 16% | 84% | 10% | 50% | 40% |
| treinamento Tendência D+3.arff | 17% | 83% | 9% | 46% | 45% |
| treinamento Tendência D+5.arff | 18% | 82% | 10% | 54% | 36% |
| treinamento Tendência D+10.arff | 17% | 83% | 11% | 50% | 40% |
| treinamento Tendência D+15.arff | 21% | 79% | 10% | 49% | 41% |
| treinamento Variação D+1.arff | 18% | 82% | 9% | 50% | 41% |
| treinamento Variação D+2.arff | 17% | 83% | 10% | 47% | 43% |
| treinamento Variação D+3.arff | 19% | 81% | 7% | 47% | 46% |
| treinamento Variação D+5.arff | 16% | 84% | 8% | 50% | 42% |
| treinamento Variação D+10.arff | 23% | 77% | 8% | 50% | 41% |
| treinamento Variação D+15.arff | 41% | 59% | 16% | 44% | 40% |
| **Médias** | 32% | 68% | 22% | 52% | 26% |

**Fonte:** Do autor.

Outro fator importante a ser observado e que influenciou na seleção do algoritmo foi a geração de atributos não classificados pelo algoritmo ID3. Isto acontece devido a poda do algoritmo funcionar de forma ineficaz no contexto, criando árvores grandes, com muitas ramificações e repetindo atributos.

Depois de selecionado o algoritmo de classificação J48, foi possível analisar os modos de testes disponibilizados pelo software de classificação. Os arquivos acima citados estavam divididos em um arquivo de teste, com 30% dos registros e um arquivo de treinamento, contendo 70% dos registros, escolhidos de forma aleatória. Para os modos de teste de *Use training set, Cross-Validation e Percentage Split* estes arquivos foram unificados, pois o próprio weka seleciona os registros de teste.

A união dos arquivos de testes e treinamento foi necessária porque todos os modos de teste do Weka, exceto o *Suppied Test Set* já fazem de forma automática a separação do percentual de teste conforme parâmetros pré-definidos, considerando como base o arquivo de treinamento. Logo, para comparar a interferência do modo de teste nos algoritmos de classificação o número de registros considerados em todos os modos deveria ser o mesmo. Sendo assim, o modo *Suppied Test Set* considerou cerca de 1250 registros em dois arquivos distintos, um de teste com 30% dos registros e outro de treinamento, com 70% dos registros, separados de forma aleatória. Já os modos *Use training set, Cross-Validation e Percentage Split* foram avaliados com um único arquivo para treinamento e teste, com os mesmos 1250 registros, de onde de forma automática o software seleciona registros e efetua a validação do(s) padrão(ões) identificado(s).

Como mostrado na tabela a seguir, o *Use training set* se destaca na média de classificação correta dos atributos com 61%, contra cerca de 32% dos demais modos de testes. Entretanto o uso do próprio arquivo de treinamento para teste do modelo de mineração pode trazer interferências nos resultados. Por este motivo, optou-se por utilizar o modo *Supplied test set*, que utiliza um arquivo de teste com 30% do total de registros, separado do arquivo de treinamento, evitando assim interferências nos testes.

**Tabela 5.1:** Comparação entre os modos de testes dos algoritmos.

| Arquivo | Modo de Teste | | | |
|---|---|---|---|---|
| | Supplied test Set | Use trainning set | Cross-Validation | Percentage Split |
| | Correto | Correto | Correto | Correto |
| TREMaiorOuMenor0D+5.arff | 72,33% | 76,08% | 72,00% | 72,00% |
| TREMaiorOuMenor0D+2.arff | 75,81% | 76,08% | 76,08% | 77,00% |
| treinamento0a3.arff | 27,08% | 57,20% | 31,24% | 31,00% |
| treinamento0a3DiaPost.arff | 48,61% | 62,00% | 47,05% | 48,00% |
| treinamento0a3DiaPost_Sem IFR.arff | 49,07% | 59,00% | 47,21% | 45,00% |
| treinamento0a5.arff | 45,83% | 59,17% | 48,23% | 45,00% |
| treinamento0a5diaPost.arff | 56,94% | 65,00% | 52,95% | 50,00% |
| Treinamento5.arff | 58,10% | 67,98% | 49,65% | 48,00% |
| Treinamento.arff | 15,74% | 60,90% | 16,60% | 18,00% |
| Treinamento_FaixasReclass.arff | 40,74% | 51,38% | 40,76% | 44,00% |
| Treinamento_FaixasReclassDiaPosterior.arff | 48,38% | 61,61% | 52,71% | 51,00% |
| Treinamento_PETR.arff | 17,82% | 30,84% | 19,67% | 16,00% |
| Treinamento_SemIFR.arff | 13,43% | 59,64% | 18,49% | 18,00% |
| Treinamento_SemIfrDiario.arff | 19,68% | 52,16% | 17,23% | 20,00% |
| Treinamento_VolReclass.arff | 17,13% | 57,28% | 19,28% | 18,00% |
| TReVarD+2.arff | 28,37% | 63,49% | 29,35% | 29,00% |
| TReVarD+5.arff | 36,28% | 63,97% | 35,41% | 37,00% |
| TreHoje | 30,23% | 62,53% | 35,57% | 37,00% |
| TreHoje+5 | 34,03% | 55,19% | 33,62% | 33,00% |
| TreHoje+10 | 33,33% | 59,63% | 35,35% | 31,00% |
| TreHoje+15 | 31,78% | 64,11% | 32,85% | 31,00% |
| TreHoje+20 | 31,76% | 67,34% | 33,15% | 34,00% |
| Treinamento Sexta.arff | 35,56% | 67,45% | 37,65% | 37,00% |
| Treinamento_Quinta.arff | 23,75% | 58,00% | 26,56% | 25,00% |
| Treinamento_Quarta.arff | 40,22% | 65,53% | 36,36% | 51,00% |
| Treinamento_Terça.arff | 32,97% | 62,99% | 35,43% | 25,00% |
| Treinamento_Segunda.arff | 42,17% | 68,46% | 39,42% | 41,00% |
| Treinamento_SegundaSemD+5.arff | 39,76% | 71,37% | 45,64% | 30,00% |
| treinamento Tendência D+1.arff | 18,31% | 55,08% | 18,36% | 18,00% |
| treinamento Tendência D+2.arff | 16,24% | 58,68% | 16,72% | 18,75% |
| treinamento Tendência D+3.arff | 16,75% | 62,27% | 17,05% | 17,00% |
| treinamento Tendência D+5.arff | 17,77% | 61,66% | 18,66% | 17,00% |
| treinamento Tendência D+10.arff | 17,27% | 61,75% | 15,63% | 18,00% |
| treinamento Tendência D+15.arff | 21,36% | 63,03% | 14,82% | 18,00% |
| treinamento Variação D+1.arff | 17,52% | 54,25% | 18,90% | 20,00% |
| treinamento Variação D+2.arff | 17,10% | 58,23% | 15,45% | 18,00% |
| treinamento Variação D+3.arff | 19,44% | 60,33% | 18,85% | 19,00% |

| Arquivo | Modo de Teste | | | |
|---|---|---|---|---|
| | Supplied test Set | Use trainning set | Cross-Validation | Percentage Split |
| | Correto | Correto | Correto | Correto |
| treinamento Variação D+5.arff | 15,76% | 57,98% | 14,93% | 13,00% |
| treinamento Variação D+10.arff | 22,75% | 56,30% | 25,46% | 28,00% |
| treinamento Variação D+15.arff | 40,57% | 58,52% | 41,16% | 38,00% |
| **Médias** | 32% | 61% | 33% | 32% |

**Fonte:** Do autor.

Tomando como base o teste entre os algoritmos ID3 e J48 realizado anteriormente, foram selecionados os arquivos que se destacaram nas classificações corretas para definir os melhores parâmetros a serem utilizados neste algoritmo. Logo, os arquivos TREMaiorOuMenor0D+5.arff e TREMaiorOuMenor0D+2.arff, respectivamente com 72% e 76% de classificações corretas são utilizados para esta definição, conforme mostrado na tabela 5.1. Os testes consistiram em avaliar todos os arquivos no modo de teste *Supplied test set*, e buscar os melhores valores para cada parâmetro. Como critério de avaliação, foi considerado a média das classificações corretas dos dois arquivos testados.

Os arquivos TREMaiorOuMenor0D+5.arff e TREMaiorOuMenor0D+2.arff foram preparados com valores de faixas separados por variações positivas e negativas para todos os atributos de todos ativos estudados. O arquivo TREMaiorOuMenor0D+5.arff busca a relação do valor de fechamento da ação da Petr4 da data atual com 5 dias à frente e o arquivo e TREMaiorOuMenor0D+2.arff busca a relação entre este atributo com 2 dias à frente.

Os parâmetros de ajuste do algoritmo J48 foram então estudados de forma isolada. Visando obter o melhor valor para o parâmetro fator de confiança (*Confidence Factor)*, foram comparadas os testes 1,2,3 e 4, mantendo os demais atributos estáveis:

**Tabela 5.2:** Comparação de testes variando o valor do parâmetro *Confidence Factor*.

| **Parâmetro:** | **Teste01** | **Teste02** | **Teste03** | **Teste04** |
|---|---|---|---|---|
| binary Splits | False | False | False | False |
| Confidence Factor | 0,25 | 1 | 0,5 | 0,1 |
| debug | False | False | False | False |
| minNumObj | 2 | 2 | 2 | 2 |
| NumFolds | 3 | 3 | 3 | 3 |
| ReducedErrorPruning | False | False | False | False |
| SaveInstanceData | False | False | False | False |
| Seed | 1 | 1 | 1 | 1 |
| SubtreeRaising | True | True | True | True |
| unpruned | False | False | False | False |
| UseLaplace | False | False | False | False |
| TreMaiorOuMenor0D+2 | 75,81% | 65,58% | 73,25% | 75,81% |

| | | | | |
|---|---|---|---|---|
| TreMaiorOuMenor0D+5 | 72,32% | 65,58% | 66,51% | 73,72% |
| Média | 74,07% | 65,58% | 69,88% | 74,77% |

**Fonte:** Do autor.

Com estas comparações, fica definido que o melhor valor para o atributo avaliado é 0,1 pois o teste 04 mostrou a melhor média de classificação com 74,77% de classificações corretas. Cabe salientar que o ganho nos resultados foi notado apenas no estudo com o arquivo TREMAiorOuMenor0D+5, já que o arquivo TREMAiorOuMenor0D+2 teve o percentual de acerto máximo repetido nos testes 01 e 04. Nota-se também a interferência do fator de confiança na poda da árvore, conforme a figura a seguir:

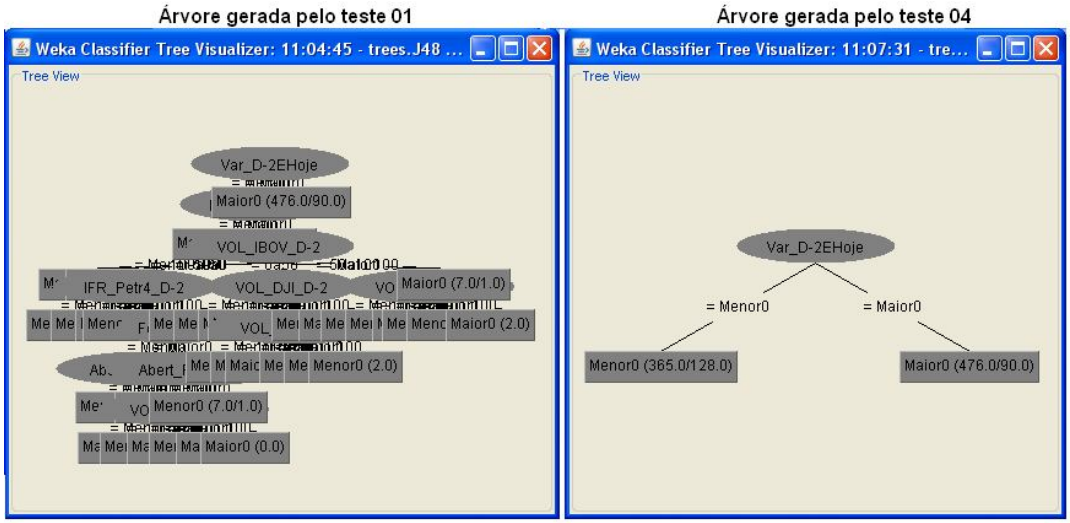


**Figura 5.1:** Árvores geradas nos testes 01 e 04 para o arquivo TREMaiorOuMenor0D+5.
**Fonte:** Do autor.

Conforme citado por Monteiro (2005), o valor do fator de confiança (*Confidence Factor)* é reduzido para forçar uma poda "maior" e obter um modelo mais genérico. Mas nos casos em que os dados de teste não variam significativamente com relação aos dados de treino, o fator de confiança pode ser aumentado de forma a obter-se um modelo com uma árvore mais detalhada. No presente estudo, visando obter uma estrutura mais enxuta foi utilizado o menor valor.

Para verificar o valor mais adequado ao parâmetro *Binary Splits*, foram comparados os testes de números 01 e 05, mantendo os demais atributos inalterados. Os resultados obtidos são mostrados na tabela a seguir:

**Tabela 5.3:** Comparação de testes variando o valor de *Binary Splits*

| **Parâmetro:** | **teste01** | **teste05** |
|---|---|---|
| binary Splits | False | True |
| Confidence Factor | 0,25 | 0,25 |
| debug | False | False |
| minNumObj | 2 | 2 |
| NumFolds | 3 | 3 |
| ReducedErrorPruning | False | False |
| SaveInstanceData | False | False |
| Seed | 1 | 1 |
| SubtreeRaising | True | True |
| unpruned | False | False |
| UseLaplace | False | False |
| TreMaiorOuMenor0D+2 | 75,81% | 75,12% |
| TreMaiorOuMenor0D+5 | 72,32% | 72,56% |
| Média | 74,07% | 73,84% |

**Fonte:** Do autor

O parâmetro *Bynary Splits* como *False* indicou uma média de classificações corretas de 74,07% contra 73,84% de classificações corretas quando o parâmetro foi definido como *True*. Desta forma a melhor definição para o parâmetro estudado é *False*.

Para verificar a interferência do parâmetro *Debug*, foram realizados testes alternando o valor deste parâmetro. Entretanto, nos testes realizados não houve ganho significativo,com a mesma média de classificação correta de 74,07% para os dois arquivos abordados, o que forçou uma escolha aleatória entre os valores testados.

**Tabela 5.4:** Comparação de testes variando o parâmetro *Debug*.

| **Parâmetro:** | **teste01** | **teste06** |
|---|---|---|
| binary Splits | False | False |
| Confidence Factor | 0,25 | 0,25 |
| debug | False | True |
| minNumObj | 2 | 2 |
| NumFolds | 3 | 3 |
| ReducedErrorPruning | False | False |
| SaveInstanceData | False | False |
| Seed | 1 | 1 |
| SubtreeRaising | True | True |
| unpruned | False | False |
| UseLaplace | False | False |
| TreMaiorOuMenor0D+2 | 75,81% | 75,81% |
| TreMaiorOuMenor0D+5 | 72,32% | 72,33% |
| Média | 74,07% | 74,07% |

**Fonte:** Do autor

No atributo *MinNUmObj*, o número de instâncias por folha, foi avaliado nos testes 01,07,08 que indicaram o melhor valor para este parâmetro no teste 07, destacando a

classificação correta para 74,19% dos registros do arquivo TreMaiorOuMenor0D+5 e abaixo de 74% para os demais testes. O arquivo TreMaiorOuMenor0D+2 não sofreu interferência em seus resultados com a alteração deste parâmetro. Os testes são mostrados a seguir:

**Tabela 5.5:** Comparação de testes variando o parâmetro *MinNumObj*.

| **Parâmetro:** | **teste01** | **teste07** | **teste08** |
|---|---|---|---|
| Binary Splits | False | False | False |
| Confidence Factor | 0,25 | 0,25 | 0,25 |
| Debug | False | False | False |
| MinNumObj | 2 | 10 | 20 |
| NumFolds | 3 | 3 | 3 |
| ReducedErrorPruning | False | False | False |
| SaveInstanceData | False | False | False |
| Seed | 1 | 1 | 1 |
| SubtreeRaising | True | True | True |
| Unpruned | False | False | False |
| UseLaplace | False | False | False |
| TreMaiorOuMenor0D+2 | 75,81% | 75,81% | 75,81% |
| TreMaiorOuMenor0D+5 | 72,32% | 74,19% | 73,72% |
| Média | 74,07% | 75,00% | 74,77% |

**Fonte:** Do autor

O atributo *NumFolds* foi avaliado em três testes de números 1,9,10. Os testes mostram que a alteração do parâmetro *NumFlods* não interfere no resultado. Entretanto este fator tem interferência quando o parâmentro *ReducedErrorPrunig* é definido como *TRUE*, nos testes 11,12 e 13, apesar de indicarem resultados inferiores aos testes 01,09 e 10. Mas mesmo após os testes, os melhores resultados na média geral foram os obtidos com o parâmetro *ReducedErrorPruning* em *False*. Logo, qualquer um dos valores testados para *NumFolds* pode ser utilizado como valor ideal, onde determinou-se o valor 3 como padrão. Para o parâmetro *ReducedErrorPruning*, o melhor definição é *False*, conforme mostrado na comparação dos testes 10 e 11.

**Tabela 5.6:** Comparação de testes variando o parâmetro *NumFolds*.

| **Parâmetro:** | **teste01** | **teste09** | **teste10** | **teste11** | **teste12** | **teste13** |
|---|---|---|---|---|---|---|
| Binary Splits | False | False | False | False | False | False |
| Confidence Factor | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 |
| Debug | False | False | False | False | False | False |
| MinNumObj | 2 | 2 | 2 | 2 | 2 | 2 |
| NumFolds | 3 | 10 | 20 | 20 | 10 | 3 |
| ReducedErrorPruning | False | False | False | True | True | True |
| SaveInstanceData | False | False | False | False | False | False |
| Seed | 1 | 1 | 1 | 1 | 1 | 1 |
| SubtreeRaising | True | True | True | True | True | True |
| Unpruned | False | False | False | False | False | False |
| UseLaplace | False | False | False | False | False | False |

| Parâmetro: | teste01 | teste09 | teste10 | teste11 | teste12 | teste13 |
|---|---|---|---|---|---|---|
| TreMaiorOuMenor0D+2 | 75,81% | 75,81% | 75,81% | 72,32% | 74,19% | 73,49% |
| TreMaiorOuMenor0D+5 | 72,32% | 72,33% | 72,33% | 73,26% | 71,40% | 70,93% |
| Média | 74,07% | 74,07% | 74,07% | 72,79% | 72,79% | 72,21% |

**Fonte:** Do autor

O parâmetro *SaveInstanceData* que define o salvamento ou não dos resultados não foi testado porque não interfere no resultado das classificações.

Os Testes 1,14,15 avaliaram o parâmetro *seed*, que com valores alternados não apresentaram ganho nos resultados de classificação, mantendo em todos o testes a mesma média de 74,07% de classificações corretas. O mesmo ocorreu na comparação dos testes 01 e 16 que avaliaram o parâmetro *SubtreeRaising*, como *True* ou *False*, também sem alterações nos resultados. Os testes acima citados são mostrados na tabela a seguir:

**Tabela 5.7:** Comparação de testes variando os parâmetros Seed e SubTreeRaising.

| Parâmetro: | teste01 | teste14 | teste15 | teste16 |
|---|---|---|---|---|
| Binary Splits | False | False | False | False |
| Confidence Factor | 0,25 | 0,25 | 0,25 | 0,25 |
| Debug | False | False | False | False |
| MinNumObj | 2 | 2 | 2 | 2 |
| NumFolds | 3 | 3 | 3 | 3 |
| ReducedErrorPruning | False | False | False | False |
| SaveInstanceData | False | False | False | False |
| Seed | 1 | 5 | 10 | 1 |
| SubtreeRaising | True | True | True | False |
| Unpruned | False | False | False | False |
| UseLaplace | False | False | False | False |
| TreMaiorOuMenor0D+2 | 75,81% | 75,81% | 75,81% | 75,81% |
| TreMaiorOuMenor0D+5 | 72,32% | 72,32% | 72,32% | 72,32% |
| Média | 74,07% | 74,07% | 74,07% | 74,07% |

**Fonte:** Do autor

A comparação dos testes 01 e 17 avaliaram o parâmetro *Unpruned* , o que gerou além de resultados piores, uma árvore gigantesca, como mostrado na figura 5.2 Logo foi alterado o parâmetro *NumFolds* no teste 18. Neste teste foram combinados os parâmetros *Unpruned* e *MinNumFolds* com novos valores, sem ganhos na classificação média com relação aos testes 01 e 17. Os testes 19 e 20 tinham como objetivo comparar os resultados da combinação do parâmetro *MinNumObjs* e *Unpruned* com outros valores relacionados ao teste 01 e 17. A classificação média destes testes foi de aproximadamente 71%, superior aos

65,81% apresentados no teste 17, e geraram uma árvore mais organizada. Entretanto a classificação média continuou inferior aos 74,07% do teste 01.

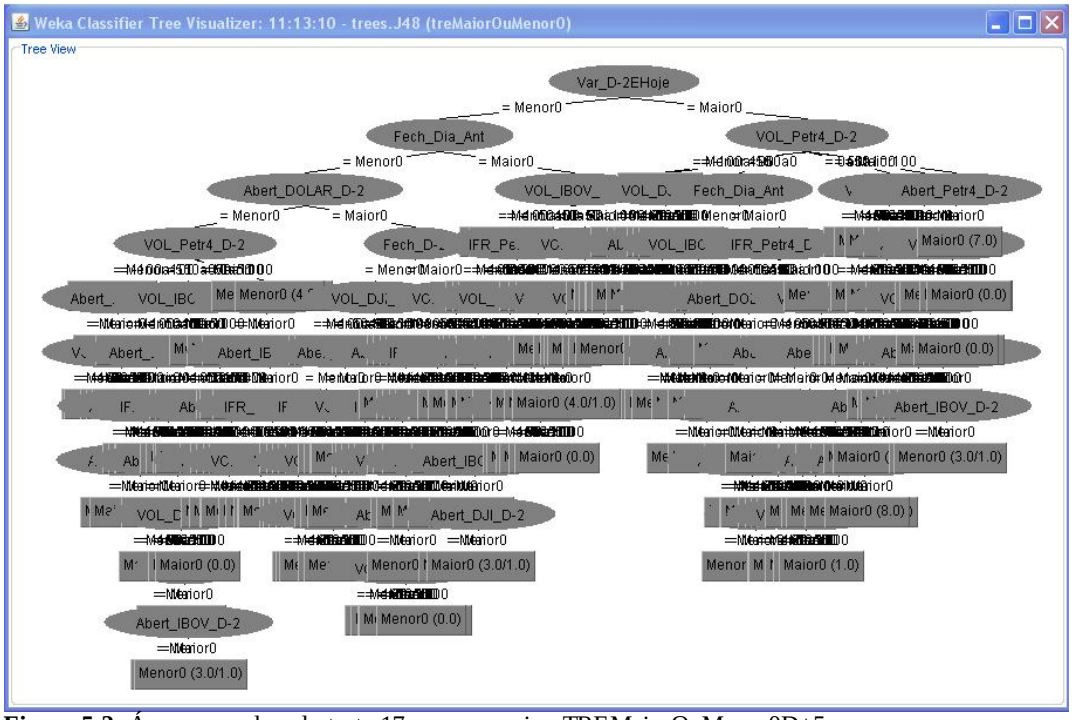


**Figura 5.2:** Árvore gerada pelo teste 17 para o arquivo TREMaiorOuMenor0D+5
**Fonte:** Do autor

**Tabela 5.8:** Comparação dos testes 1,17,18,19 e 20.

| Parâmetro: | teste01 | teste17 | teste18 | Teste19 | teste20 |
|---|---|---|---|---|---|
| Binary Splits | False | False | False | False | False |
| Confidence Factor | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 |
| Debug | False | False | False | False | False |
| MinNumObj | 2 | 2 | 2 | 10 | 20 |
| NumFolds | 3 | 3 | 10 | 3 | 3 |
| ReducedErrorPruning | False | False | False | False | False |
| SaveInstanceData | False | False | False | False | False |
| Seed | 1 | 1 | 1 | 1 | 1 |
| SubtreeRaising | True | True | True | True | True |
| Unpruned | False | True | True | True | True |
| UseLaplace | False | False | False | False | False |
| TreMaiorOuMenor0D+2 | 75,81% | 66,05% | 66,05% | 71,86% | 74,18% |
| TreMaiorOuMenor0D+5 | 72,32% | 65,58% | 65,58% | 70,23% | 71,86% |
| Média | 74,07% | 65,81% | 65,81% | 71,05% | 73,02% |

**Fonte:** Do autor

O teste 21 buscou avaliar o parâmetro *UseLaplace* com relação aos resultados do teste 01, mas a média de acerto dos dois arquivos foi a mesma para os dois testes realizados, o que levou a escolha do parâmetro *False* adotado no teste 01 através de uma escolha aleatória.

**Tabela 5.9:** Resultados dos testes com alteração do parâmentro UseLaplace.

| Parâmetro: | teste01 | teste21 |
|---|---|---|
| Binary Splits | False | False |
| Confidence Factor | 0,25 | 0,25 |
| Debug | False | False |
| MinNumObj | 2 | 2 |
| NumFolds | 3 | 3 |
| ReducedErrorPruning | False | False |
| SaveInstanceData | False | False |
| Seed | 1 | 1 |
| SubtreeRaising | True | True |
| Unpruned | False | False |
| UseLaplace | False | True |
| TreMaiorOuMenor0D+2 | 75,81% | 75,81% |
| TreMaiorOuMenor0D+5 | 72,32% | 72,33% |
| Média | 74,07% | 74,07% |

**Fonte:** Do autor

Baseados nos testes realizados foram identificados os melhores valores para cada um dos parâmetros do algoritmo J48. Com isso, foi realizado o teste 22, utilizando os melhores valores de cada parâmetro nos arquivos testados, conforme mostrado a seguir. Nesta tabela são observadas as diferenças entre o teste de melhor classificação (teste 07) e o teste com os melhores parâmetros (teste 22).

**Tabela 5.10:** Teste com os melhores parâmetros e melhor classificação.

| Parâmetro: | teste07 | teste22 |
|---|---|---|
| Binary Splits | False | False |
| Confidence Factor | 0,25 | 0.1 |
| Debug | False | False |
| MinNumObj | 10 | 10 |
| NumFolds | 3 | 3 |
| ReducedErrorPruning | False | False |
| SaveInstanceData | False | False |
| Seed | 1 | 1 |
| SubtreeRaising | True | True |
| Unpruned | False | False |
| UseLaplace | False | False |
| TreMaiorOuMenor0D+2 | 75,81% | 75,81% |
| TreMaiorOuMenor0D+5 | 74,19% | 73,72% |
| Média | 75,00% | 74,77% |

**Fonte:** Do autor

No teste 22, foi obtida uma média de classificação geral de 74,77% para os dois arquivos testados, onde o arquivo TREMaiorOuMenor0D+5 apresentou 73,72% de classificações corretas e o arquivo TREMaiorOuMenor0D+2 apresentou 75,81%.

Entretanto, o teste 07 apresentou um valor de classificação médio para os dois arquivos superior ao teste com os melhores valores de cada atributo (Teste 22), contrariando a metodologia abordada. A diferença entre os testes 07 e 22 está no valor do *Confidence Factor*, que no teste 07 é de 0,25 e o melhor valor indicado pelos testes para este atributo é de 0,1, adotado no teste 22. A diferença em classificações corretas entre estes dois arquivos mostrou uma superioridade do teste 07 de 0,23% com relação ao teste 22. Com isso, é possível concluir que os parâmetros sofrem interferência conforme os valores dos outros parâmetros adotados para o algoritmo. Todos os testes realizados estão disponíveis no anexo A.

Os valores parametrizados no teste 07 foram então definidos como os mais adequados ao ambiente e metodologia abordados, apresentando as seguintes árvores para os arquivos TREMAiorOuMenor0D+2 e TREMAiorOuMenor0D+5, Respectivamente:

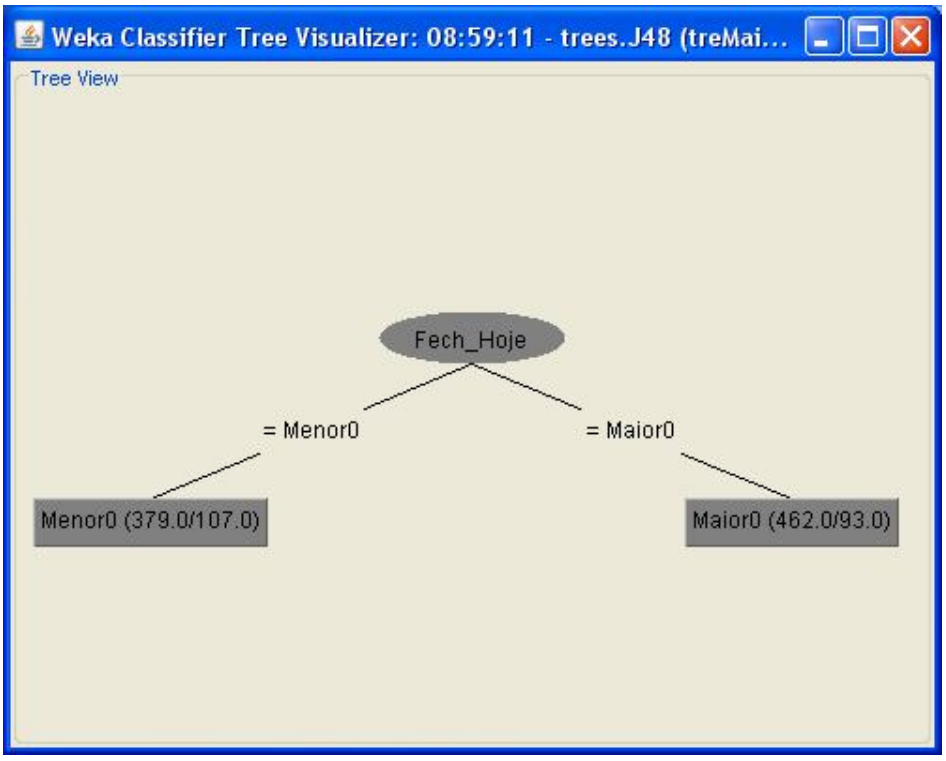


**Figura 5.3**: Árvore gerada pelo Teste 07 para o arquivo TREMaiorOuMenor0D+2 .
**Fonte:** Do autor

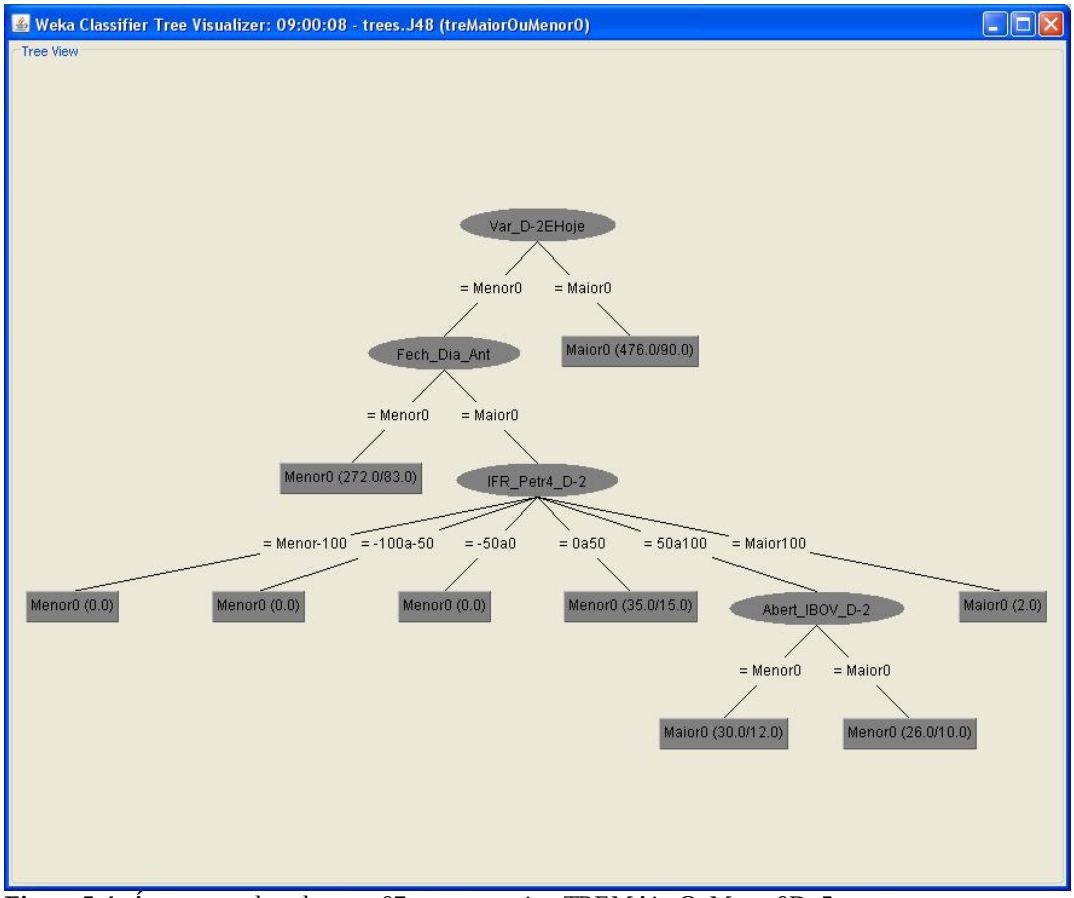


**Figura 5.4:** Árvore gerada pelo teste 07 para o arquivo TREMAiorOuMenor0D+5.
**Fonte:** Do autor.

A estrutura da árvore do arquivo TREMaiorOuMenorD+2 é simples, conforme mostrado na figura 5.3. Já a árvore gerada pelo arquivo TREMaiorOuMenorD+5 possui uma estrutura complexa, como pode-se visualizar na árvore gerada mostrada na figura 5.4. A imagem desta árvore é confusa devido a grande quantidade de folhas geradas.

Cabe salientar que os demais arquivos, apesar de não apresentarem um percentual alto na classificação dos modelos, podem internamente apresentar relações úteis ao escopo do trabalho, como regras com alto grau de classificações corretas. Para explorar esta possibilidade, todas as regras geradas por estes arquivos através das árvores de decisão, foram listadas e devidamente identificadas, ou seja, todos os ramos de cada folha da árvore tiveram seu caminho descrito.

Para exemplificar, abaixo são mostradas as regras geradas pelo software Weka:

```
Var_D-2EHoje = Menor0
|   Fech_Dia_Ant = Menor0: Menor0 (272.0/83.0)
|   Fech_Dia_Ant = Maior0
|   |   IFR_Petr4_D-2 = Menor-100: Menor0 (0.0)
|   |   IFR_Petr4_D-2 = -100a-50: Menor0 (0.0)
|   |   IFR_Petr4_D-2 = -50a0: Menor0 (0.0)
|   |   IFR_Petr4_D-2 = 0a50: Menor0 (35.0/15.0)
|   |   IFR_Petr4_D-2 = Maior100: Maior0 (2.0)
Var_D-2EHoje = Maior0: Maior0 (476.0/90.0)
```

**Quadro: 5.1** Regras geradas pelo software Weka.
**Fonte:** Do autor.

As regras mostradas acima, que buscam definir o padrão para VarD+5 podem ser descritas utilizando condições e operadores condicionais (Se/Então), da seguinte forma:

```
Se Var_D-2EHoje = Menor0 e Fech_Dia_Ant = Menor0 Então VarD-2eD+3 = Menor0 (272.0/83.0)
Se Var_D-2EHoje = Menor0 e Fech_Dia_Ant = Maior0 e IFR_Petr4_D-2 = Menor-100 Então VarD-2eD+3 = Menor0 (0.0)
Se Var_D-2EHoje = Menor0 e Fech_Dia_Ant = Maior0 e IFR_Petr4_D-2 = -100a-50 Então VarD-2eD+3 = Menor0 (0.0)
Se Var_D-2EHoje = Menor0 e Fech_Dia_Ant = Maior0 e IFR_Petr4_D-2 = -50a0 Então VarD-2eD+3 = Menor0 (0.0)
Se Var_D-2EHoje = Menor0 e Fech_Dia_Ant = Maior0 e IFR_Petr4_D-2 = 0a50: Então VarD-2eD+3 = Menor0 (35.0/15.0)
Se Var_D-2EHoje = Maior0 Então VarD-2eD+3= Maior0 (476.0/90.0)
```

**Quadro: 5.2** Regras geradas com seu caminho descrito.
**Fonte:** Do autor.

Desta forma, e através da identificação dos resultados certos ou errados para cada regra, exibidos entre parênteses, é possível a classificação das melhores regras de todos os arquivos. No total, dos 40 arquivos avaliados, foram geradas mais de 20.000 regras. Destas regras, muitas com baixíssimos níveis de acertos ou de amostras enquadradas no perfil. Além disso, diversas regras se anularam, pois apesar de originarem arquivos diferentes, possuíam o mesmo sentido prático.

Para delimitar o escopo da abordagem de regras internas dos arquivos, foi determinado empiricamente um ponto de corte, com no mínimo 50 registros enquadrados e 70% de classificações corretas. Neste padrão ainda foram posteriormente eliminadas as regras repetidas, restando apenas 17, listadas no anexo B. Estas regras podem ser utilizadas como base para a formação de um sistema especialista, conforme citado por Rezende (2005), monitorando o mercado financeiro e avisando o usuário quando um perfil for encontrado. Para este estudo não foi utilizado a opção "*minNumObj*" que determina o número mínimo de registros por folha, pois podem haver regras com alto percentual de acerto com valores menores de registros por folhas. Como este não é o foco deste estudo, esta possibilidade torna-se um assunto para trabalhos futuros.

## 5.2 Treinamento e testes com redes neurais

Com os arquivos TREMaiorOuMenor0D+5 e TREMaiorOuMenor0D+2 identificados como destaque do algoritmo de classificação J48, foram colocados em testes com o uso de redes neurais no software Weka. Para tanto, foi utilizado a rede *multilayer perceptron*, presente no pacote de funções, que utiliza como função de ativação a função Sigmóide. Como esta técnica pode apresentar características de classificação diferentes da técnica de classificação de árvores de decisão, todos os arquivos testados anteriormente foram novamente testados com a rede *multilayer perceptron*. Os resultados mostram equivalência entre as duas técnicas, confirmando as estimativas das árvores de decisão.

**Tabela 5.11:** Resultados das classificações de RNA e árvores de decisão.

| Arquivo | J48 | MLP |
|---|---|---|
| TREMaiorOuMenor0D+5.arff | 74,19% | 62,79% |
| TREMaiorOuMenor0D+2.arff | 75,81% | 66,74% |
| treinamento0a3.arff | 33,79% | 25,00% |
| treinamento0a3DiaPost.arff | 50,00% | 44,44% |
| treinamento0a3DiaPost_Sem IFR.arff | 50,00% | 44,44% |
| treinamento0a5.arff | 48,14% | 39,58% |
| treinamento0a5diaPost.arff | 57,64% | 52,78% |
| Treinamento5.arff | 54,86% | 50,69% |
| Treinamento.arff | 23,38% | 15,05% |
| Treinamento_FaixasReclass.arff | 41,90% | 33,10% |
| Treinamento_FaixasReclassDiaPosterior.arff | 53,47% | 48,38% |
| Treinamento_PETR.arff | 17,59% | 19,91% |
| Treinamento_SemIFR.arff | 23,61% | 15,97% |
| Treinamento_SemIfrDiario.arff | 17,36% | 15,05% |
| Treinamento_VolReclass.arff | 18,98% | 17,36% |
| TReVarD+2.arff | 29,77% | 26,05% |
| TReVarD+5.arff | 37,21% | 32,56% |
| TreHoje | 37,67% | 35,12% |
| TreHoje+5 | 40,09% | 28,21% |
| TreHoje+10 | 36,83% | 31,93% |
| TreHoje+15 | 36,92% | 28,74% |
| TreHoje+20 | 38,59% | 27,06% |
| Treinamento Sexta.arff | 41,11% | 33,33% |
| Treinamento_Quinta.arff | 26,25% | 23,75% |
| Treinamento_Quarta.arff | 36,96% | 33,70% |
| Treinamento_Terça.arff | 29,67% | 21,98% |
| Treinamento_Segunda.arff | 44,58% | 31,33% |
| Treinamento_SegundaSemD+5.arff | 39,76% | 28,92% |
| treinamento Tendência D+1.arff | 21,36% | 16,20% |
| treinamento Tendência D+2.arff | 18,59% | 14,35% |
| treinamento Tendência D+3.arff | 17,92% | 14,62% |
| treinamento Tendência D+5.arff | 18,72% | 15,40% |
| treinamento Tendência D+10.arff | 15,59% | 16,55% |
| treinamento Tendência D+15.arff | 17,96% | 16,99% |

| Arquivo | J48 | MLP |
|---|---|---|
| treinamento Variação D+1.arff | 19,39% | 16,36% |
| treinamento Variação D+2.arff | 20,37% | 16,86% |
| treinamento Variação D+3.arff | 21,32% | 15,69% |
| treinamento Variação D+5.arff | 16,00% | 13,88% |
| treinamento Variação D+10.arff | 26,30% | 21,56% |
| treinamento Variação D+15.arff | 40,33% | 29,83% |

**Fonte:** Do autor

Feito isso, os arquivos TREMaiorOuMenor0D+5 e TREMaiorOuMenor0D+2, juntamente com seus respectivos arquivos de testes são apresentados à redes com diferentes parametrizações, como tamanho do conjunto de validação, número de ciclos por época e número de épocas, visando obter a rede com maior grau de classificação. Foram criadas 39 redes com parametrizações diferentes, visando obter a melhor classificação. Como critério de classificação, foi considerada a média das classificações corretas dos dois arquivos testados, TREMaiorOuMenor0D+5 e TREMaiorOuMenor0D+2.

Cabe observar que quando o algoritmo *Multilayer Perceptron* utiliza atributos com faixas e não valores puros, cada faixa de cada valor torna-se uma entrada da rede a ser avaliada, como mostrado na figura a seguir:

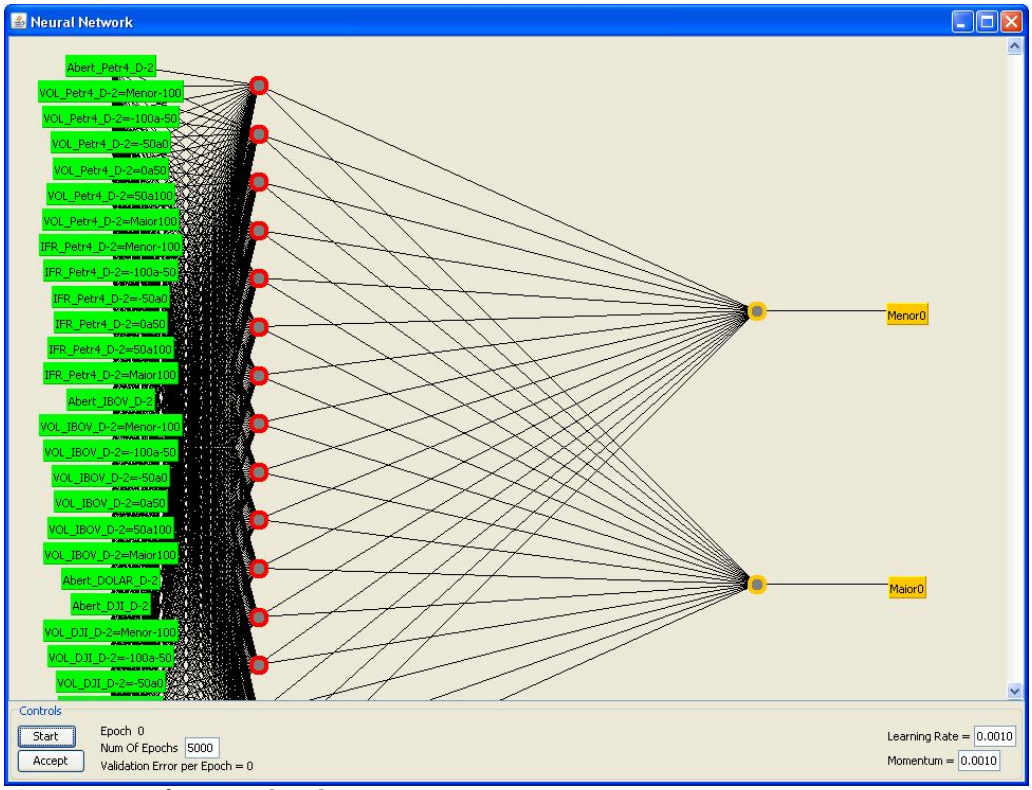


**Figura 5.5:** configuração da rede MLP para o arquivo TREMaiorOuMenor0D+5.
**Fonte:** Do autor.

A figura anterior está incompleta porque o número de entradas é muito grande e por limitações do software não foi possível visualizar os demais atributos de entrada.

Os parâmetros que mais influenciam na rede foram então estudados de forma isolada. Visando obter o melhor valor para o parâmetro *Learning Rate*, foram comparadas as redes dos testes 29, 31,34 e 35, mantendo os demais parâmetros estáveis:

**Tabela 5.12:** Comparação de testes variando o valor do *Learning Rate*

| **Parâmetros/Arquivos** | **Teste 31** | **Teste 29** | **Teste 34** | **Teste 35** |
|---|---|---|---|---|
| GUI | False | False | False | False |
| Auto Build | True | True | True | True |
| Debug | False | False | False | False |
| Decay | False | False | False | False |
| Hidden Layers | a | a | a | a |
| Learning Rate | 0.00001 | 0.001 | 0,1 | 0,05 |
| Momentum | 0,01 | 0,01 | 0,01 | 0,01 |
| NominalToinaryFilter | True | True | True | True |
| Normalize Attributes | True | True | True | True |
| Normalize Numeric Class | True | True | True | True |
| Random Seed | 0 | 0 | 0 | 0 |
| Reset | True | True | True | True |
| Training Time | 10000 | 10000 | 10000 | 10000 |
| ValidationSetSize | 20 | 20 | 20 | 20 |
| Validation Threshold | 20 | 20 | 20 | 20 |
| TREMaiorOuMenor0D+2 | 72,79% | 75,81% | 74,65% | 75,58% |
| TREMaiorOuMenor0D+5 | 60,93% | 73,95% | 73,49% | 73,95% |
| Média de Classificação : | 66,86% | 74,88% | 74,07% | 74,77% |

**Fonte:** Do autor

Com estas comparações, fica definido que o melhor valor para o item avaliado é 0,001 pois o teste 29 mostrou a melhor média de classificações corretas, com 74,88% .

Para verificar o valor mais adequado ao parâmetro *Momentum*, foram comparados os testes de números 14, 15, 9 e 33, mantendo os demais inalterados. Os resultados obtidos são mostrados a seguir:

**Tabela 5.13:** Comparação de testes variando o valor de *Momentum*.

| **Parâmetros/Arquivos** | **Teste 14** | **Teste 15** | **Teste 9** | **Teste 33** |
|---|---|---|---|---|
| GUI | False | False | False | False |
| Auto Build | True | True | True | True |
| Debug | False | False | False | False |
| Decay | False | False | False | False |
| Hidden Layers | a | a | a | a |
| Learning Rate | 0,005 | 0,005 | 0,005 | 0,005 |
| Momentum | 0,001 | 0,1 | 0,0001 | 0,00001 |

| Parâmetros/Arquivos | Teste 14 | Teste 15 | Teste 9 | Teste 33 |
|---|---|---|---|---|
| NominalToinaryFilter | True | True | True | True |
| Normalize Attributes | True | True | True | True |
| Normalize Numeric Class | True | True | True | True |
| Random Seed | 0 | 0 | 0 | 0 |
| Reset | True | True | True | True |
| Training Time | 10000 | 10000 | 10000 | 10000 |
| ValidationSetSize | 70 | 70 | 70 | 70 |
| Validation Threshold | 20 | 20 | 20 | 20 |
| TREMaiorOuMenor0D+2 | 74,88% | 74,88% | 74,88% | 74,88% |
| TREMaiorOuMenor0D+5 | 72,33% | 72,09% | 72,33% | 72,33% |
| Média de Classificação : | 73,60% | 73,49% | 73,60% | 73,60% |

**Fonte:** Do autor

Nota-se que valores abaixo de 0,1 fornecem o mesmo valor de 73,60% classificações corretas para o arquivo TREMaiorOuMenor0D+5. Com isso, o valor pode ser qualquer valor entre 0,01 e 0,00001.

Para verificar a interferência do número de ciclos de treinamentos foram comparados quatro arquivos com 500, 2000, 5000 e 10000 ciclos. Entretanto, nos testes realizados a média de classificações corretas não sofreu alterações com os diferentes testes de ciclos, mostrando um valore percentual de 71,28%, o que forçou uma escolha aleatória entre os valores testados. Foi definido que 5000 ciclos seria o melhor valor para este parâmetro.

**Tabela 5.14:** Comparação de testes variando o número de ciclos.

| Parâmetros/Arquivos | T01 | T 36 | T37 | T38 |
|---|---|---|---|---|
| GUI | False | False | False | False |
| Auto Build | True | True | True | True |
| Debug | False | False | False | False |
| Decay | False | False | False | False |
| Hidden Layers | a | a | a | a |
| Learning Rate | 0.2 | 0.2 | 0.2 | 0.2 |
| Momentum | 0.3 | 0.3 | 0.3 | 0.3 |
| NominalToinaryFilter | True | True | True | True |
| Normalize Attributes | True | True | True | True |
| Normalize Numeric Class | True | True | True | True |
| Random Seed | 0 | 0 | 0 | 0 |
| Reset | True | True | True | True |
| Training Time | 2000 | 5000 | 10000 | 500 |
| ValidationSetSize | 70 | 70 | 70 | 70 |
| Validation Threshold | 20 | 20 | 20 | 20 |
| TREMaiorOuMenor0D+2 | 72,33% | 72,33% | 72,33% | 72,33% |
| TREMaiorOuMenor0D+5 | 70,23% | 70,23% | 70,23% | 70,23% |
| Média de Classificação : | 71,28% | 71,28% | 71,28% | 71,28% |

**Fonte:** Do autor

O tamanho do set de validação ( *Validation Set Size)* foi avaliado na comparação da média das classificações corretas, observando o valor ideal em 20, conforme mostrado na tabela abaixo, onde o percentual de classificações corretas ficou em 74,87%:

**Tabela 5.15:** Comparação de testes variando o atributo *Validation Set Size.*

| **Parâmetros/Arquivos** | **Teste 04** | **Teste 05** | **Teste 06** | **Teste 07** | **Teste 08** | **Teste 22** | **Teste 23** | **Teste 24** | **Teste 25** |
|---|---|---|---|---|---|---|---|---|---|
| GUI | False | False | False | False | False | False | False | False | False |
| Auto Build | True | True | True | True | True | True | True | True | True |
| Debug | False | False | False | False | False | False | False | False | False |
| Decay | False | False | False | False | False | False | False | False | False |
| Hidden Layers | a | a | a | a | a | a | a | a | a |
| Learning Rate | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 |
| Momentum | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 |
| NominalToinaryFilter | True | True | True | True | True | True | True | True | True |
| Normalize Attributes | True | True | True | True | True | True | True | True | True |
| Normalize Numeric Class | True | True | True | True | True | True | True | True | True |
| Random Seed | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Reset | True | True | True | True | True | True | True | True | True |
| Training Time | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 |
| ValidationSetSize | 0 | 80 | 90 | 70 | 60 | 10 | 5 | 20 | 30 |
| Validation Threshold | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| TREMaiorOuMenor0D+2 | 62,56% | 74,65% | 68,37% | 75,58% | 73,72% | 75,34% | 75,58% | 75,58% | 74,41% |
| TREMaiorOuMenor0D+5 | 66,98% | 70,23% | 66,97% | 72,33% | 73,26% | 74,18% | 72,79% | 74,16% | 73,48% |
| Média de Classificação : | 64,77% | 72,44% | 67,67% | 73,95% | 73,49% | 74,76% | 74,19% | 74,87% | 73,95% |

**Fonte:** Do autor.

O parâmetro *Validation Threshold* foi avaliado em três testes de número 19,20 e 39, tendo o melhor desempenho os valores de 20 e 0, com 72,23% de classificações corretas. Logo, qualquer um destes dois valores pode ser utilizado como valor ideal.

**Tabela 5.16:** Comparação de testes variando o parâmetro *Validation Threshold.*

| **Parâmetros/Arquivos** | **Teste 39** | **Teste 19** | **Teste 20** |
|---|---|---|---|
| GUI | False | False | False |
| Auto Build | True | True | True |
| Debug | False | False | False |
| Decay | False | False | False |
| Hidden Layers | a | a | a |
| Learning Rate | 0,05 | 0,05 | 0,05 |
| Momentum | 0,1 | 0,1 | 0,1 |
| NominalToinaryFilter | True | True | True |
| Normalize Attributes | True | True | True |
| Normalize Numeric Class | True | True | True |
| Random Seed | 0 | 0 | 0 |
| Reset | True | True | True |
| Training Time | 10000 | 10000 | 10000 |
| ValidationSetSize | 80 | 80 | 80 |

| Parâmetros/Arquivos | Teste 39 | Teste 19 | Teste 20 |
|---|---|---|---|
| Validation Threshold | 0 | 20 | 70 |
| TREMaiorOuMenor0D+2 | 74,65% | 74,65% | 72,79% |
| TREMaiorOuMenor0D+5 | 70,23% | 70,23% | 68,60% |
| Média de Classificação : | 72,44% | 72,44% | 70,70% |

**Fonte:** Do autor.

Com base em todos estes testes, buscando identificar os melhores valores para os principais parâmetros a serem utilizados nos modelos de *multilayer perceptron*, foi testado um modelo com os valores ideais, no teste 40. Os itens testados receberam os seguintes valores: *Learning Rate*: 0,001, *Momentum*:0,001, *Training Time*:5000, *Validation Set Size*:20 e *Validation Threshold*: 20. Neste modelo, o arquivo TREMaiorOuMenor0D+5 apresentou 73,95% de classificações corretas e o arquivo TREMaiorOuMenor0D+2 apresentou 75,81%.

Entretanto, o teste 30 apresentou um valor de classificação médio para os dois arquivos de 75,35%, superior ao teste com os melhores valores de cada parâmetro, abordado no teste 40 com 74,88% de classificações corretas, contrariando a metodologia abordada (Tabela 5.18). A diferença entre os testes 40 e 30 está no valor do *Validation Set Size*, que no teste 30 é de 10 e o melhor valor indicado pelos testes para este item é de 20, adotado no teste 40. A diferença em classificações corretas entre estes dois valores durante o teste mostrou uma pequena superioridade de 0,47% para o valor de 20 com relação ao valor de 10. Com isso, é possível concluir que, assim como na técnica de árvores de decisão, os parâmetros das RNAs também sofrem interferência conforme os valores dos outros atributos da rede.

O teste 30 apresentou um percentual de 75,35% de classificações corretas, sendo este o melhor valor obtido. Neste modelo, o arquivo TREMaiorOuMenor0D+5 apresentou 74,88% de classificações corretas e o arquivo TREMaiorOuMenor0D+2 apresentou 75,81%.

**Tabela 5.17:** Comparação de testes 30 e 40.

| Parâmetros/Arquivos | T 30 | T 40 |
|---|---|---|
| GUI | False | False |
| Auto Build | True | True |
| Debug | False | False |
| Decay | False | False |
| Hidden Layers | a | a |
| Learning Rate | 0,001 | 0,001 |
| Momentum | 0,001 | 0,001 |
| NominalToinaryFilter | True | True |
| Normalize Attributes | True | True |
| Normalize Numeric Class | True | True |

| Parâmetros/Arquivos | T 30 | T 40 |
|---|---|---|
| Random Seed | 0 | 0 |
| Reset | True | True |
| Training Time | 5000 | 5000 |
| ValidationSetSize | 10 | 20 |
| Validation Threshold | 20 | 20 |
| TREMaiorOuMenor0D+2 | 75,81% | 75,81% |
| TREMaiorOuMenor0D+5 | 74,88% | 73,95% |
| Média de Classificação : | 75,35% | 74,88% |

**Fonte:** Do autor.

Como o fator de classificação média geral está sendo utilizado para definição do melhor modelo, ficam os valores de parâmetros do teste 30 como o modelo de RNA a ser utilizado para comparação com a técnica de árvores de decisão, já definida anteriormente. Os resultados, de todos os testes de parâmetros das RNAs realizados estão listados no anexo C.

Outros parâmetros como o número de neurônios da rede também podem ser ajustados conforme a necessidade do usuário, entretanto seguindo estudos já realizados, inclusive na previsão de valores de commodity com o uso de redes neurais (FREIMAN; PAMPLOMA, 2005), o número de neurônios da camada oculta utilizado para estudos no mercado financeiro é geralmente a metade do número de neurônios da camada de entrada. No estudo mostrado este fator pode ser ajustado através do parâmetro *hiddenlayers*, onde o valor "a" indica a metade dos elementos de entrada. Estes parâmetros não foram alterados, pois isto deixaria o escopo da definição do modelo de RNA muito amplo, interferindo no cronograma de estudo. Assim, surge como oportunidade de trabalhos futuros a exploração destes itens e de novos valores, buscando redes mais eficientes no estudo dos mercados financeiros.

## 5.3 Validação

Antes de validar os modelos gerados observou-se um erro na montagem dos atributos dos arquivos. Ambos os arquivos utilizam como atributos a cotação de abertura da Petrobrás (Abert_Petr4), do índice da Bovespa (Abert_IBOV) e Dow Jones (Abert_DJI), a cotação de abertura do Dolar Comercial (Abert_DOLAR), o volume de negociações da Petrobrás (VOL_Petr4), da Bovespa (VOL_IBOV ), do Dow Jones (VOL_DJI) , o IFR da Petrobrás

(IFR_Petr4), além dos valores de fechamento do dia e do dia posterior para verificar relações com a variação entre o preço de fechamento após dois ou cinco dias.

Como os arquivos foram montados com base em dados passados, estavam disponíveis tanto os dados do dia posterior, quanto os dados de dois e cinco dias posteriores para cada registro. Por exemplo, no arquivo TREMaiorOuMenor0D+2, foram utilizados dados do dia posterior para serem relacionados com os dados de dois dias à frente. Já no arquivo TREMaiorOuMenor0D+5 foram utilizados além dos dados do dia posterior os dados de dois dias à frente aos dados diários para relação com os dados de cinco dias a frente. Na realidade, o investidor não terá estes dados para a realização do estudo e atuação no mercado financeiro.

No primeiro momento, este fator tornou-se um erro que forçaria a realização de todos os treinamentos novamente. Para estudar estas interferências foram realizados testes, retirando os atributos intermediários entre os valores diários e o atributo alvo da classificação (dois ou cinco dias a frente), como mostrado no quadro a seguir.

**Configuração com atributos intermediários:**
@attribute Abert_Petr4
@attribute VOL_Petr4
@attribute IFR_Petr4
@attribute Abert_IBOV
@attribute VOL_IBOV
@attribute Abert_DOLAR
@attribute Abert_DJI
@attribute VOL_DJI
@attribute Fechamento
@attribute FechamentoPost
@attribute VarD+2
@attribute VarD+5

**Configuração sem os atributos intermediários:**
@attribute Abert_Petr4
@attribute VOL_Petr4
@attribute IFR_Petr4
@attribute Abert_IBOV
@attribute VOL_IBOV
@attribute Abert_DOLAR
@attribute Abert_DJI
@attribute VOL_DJI
@attribute Fechamento
@attribute VarD+5

**Quadro 5.3:** Exemplo de alteração de Layout dos atributos para arquivo TREMaiorOuMenor0D+5.
**Fonte:** Do autor.

Entretanto, os resultados não foram similares aos obtidos anteriormente. No caso do arquivo TREMaiorOuMenor0D+5 o resultado de classificações corretas com o uso do

algoritmo J48 o resultado passou de 74% para 61% , da mesma forma, no arquivo TREMaiorOuMenor0D+2 com o uso do algoritmo J48 o resultado passou de 76% para 60%. Nos testes com o uso de *Multilayer Perceptron*, o arquivo TREMaiorOuMenor0D+5 mostrou uma redução no percentual de classificações corretas de 74% para 52% , da mesma forma, no arquivo TREMaiorOuMenor0D+2 teve redução nas classificações corretas, passando de 76% para 50%.

**Tabela 5.18:** Resultados dos arquivos com e sem os atributos intermediários.

| Arquivo | Com atributos intermediários | | Sem atributos intermediários | |
|---|---|---|---|---|
| | J48 | *Multilayer Perceptron* | J48 | *Multilayer Perceptron* |
| TREMaiorOuMenor0D+5.arff | 74% | 74% | 61% | 52% |
| TREMaiorOuMenor0D+2.arff | 76% | 76% | 60% | 50% |

**Fonte:** Do autor.

Com estas informações, foram efetuados novos testes buscando a relação entre os dados do dia com o fechamento do dia posterior no arquivo TREMaiorOuMenor0D+2, visando identificar um padrão de relacionamento que se destacasse entre estes dados, conforme mostrado na figura a seguir:

```
Atributo alvo original para classificação do arquivo TREMaiorOuMenor0D+2:
@attribute Abert_Petr4
@attribute VOL_Petr4
@attribute IFR_Petr4
@attribute Abert_IBOV
@attribute VOL_IBOV
@attribute Abert_DOLAR
@attribute Abert_DJI
@attribute VOL_DJI
@attribute Fechamento
@attribute FechamentoPost
@attribute VarD+2

Novo atributo alvo para classificação do arquivo TREMaiorOuMenor0D+2:
@attribute Abert_Petr4
@attribute VOL_Petr4
@attribute IFR_Petr4
@attribute Abert_IBOV
@attribute VOL_IBOV
@attribute Abert_DOLAR
@attribute Abert_DJI
@attribute VOL_DJI
@attribute Fechamento
@attribute FechamentoPost

* Atributo alvo do arquivo
```

**Quadro 5.4:** Novo teste: Alteração dos atributos alvo de classificação.
**Fonte:** Do autor.

Porém, os resultados com o uso do algoritmo J48 foram de 50%, inferiores aos 76% do modelo que leva em consideração mais de um dia nos atributos. Já com o uso do Multilayer Perceptron, os resultados foram de 49% contra os 76% do modelo anterior. Desta forma, conclui-se que os dados de dias anteriores a data atual possuem um padrão de relação com os dados de datas futuras mais eficientes que a relação entre o dia atual e os dias futuros. Logo, os modelos foram mantidos, e o padrão com maior percentual de classificação foi aceito como o mais adequado para a validação dos modelos.

Para manter este padrão de atributos nos arquivos, o problema de disponibilidade de dados dos atributos que consideravam os dias posteriores persistia. A solução foi realocar o Período de amostragem no tempo, conforme mostrado na figura a seguir:

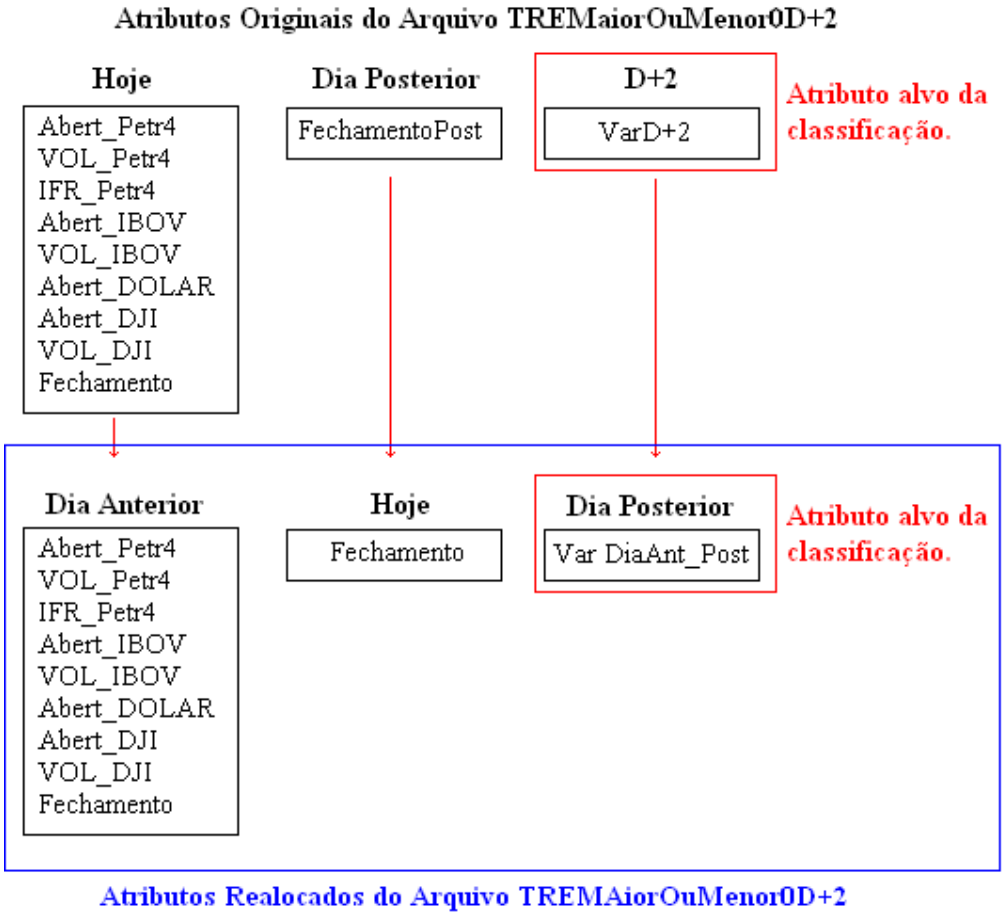


**Figura 5.6:** Alterações dos atributos do arquivo TREMaiorOuMenor0D+2.
**Fonte:** Do autor.

De forma simplificada, o arquivo TREMaiorOuMenor0D+2 relaciona os dados diários do dia anterior, com os dados do dia atual e os dados do dia seguinte, ou seja, a referência para a variação são os dados diários do dia anterior à data atual.

No caso dos modelos utilizados no arquivo TREMaiorOuMenor0D+5, os atributos abordados são os atributos diários de dois dias passados, com os dados do dia anterior e os dados do dia atual relacionando com dados de três dias posteriores. A referência, neste caso, são os dados diários de dois dias anteriores à data atual, conforme mostrado na figura a seguir:

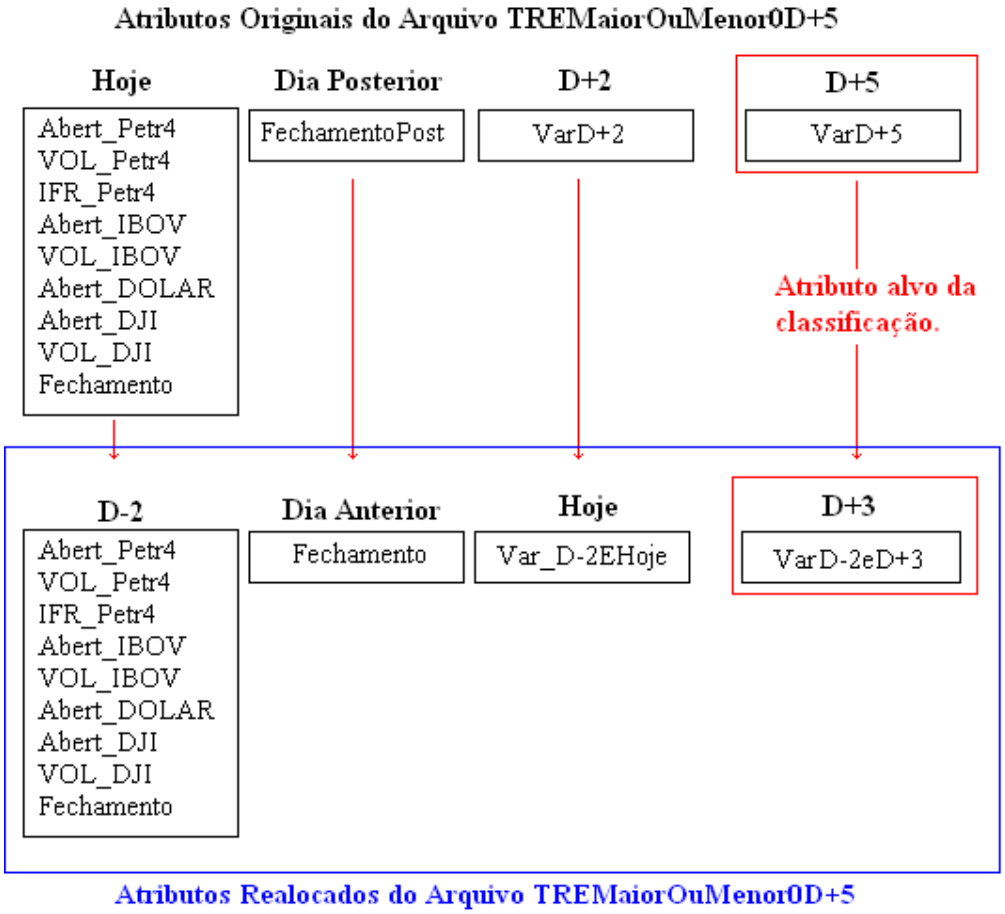


**Figura 5.7:** Alterações dos atributos do arquivo TREMaiorOuMenor0D+5.
**Fonte:** Do autor.

A validação dos modelos consiste na aplicação dos modelos encontrados em um ambiente de teste. Para a validação dos modelos no mercado financeiro da Bovespa, mais especificamente com o uso dos atributos da ação da Petrobrás, índice Bovespa, Dólar Comercial e índice Dow Jones, foram coletados os dados envolvidos no estudo dos meses de julho, agosto e setembro de 2008. Estes dados foram pré-processados, passando por etapas de limpeza, tratamento e normalização, como explicadas anteriormente, antes da aplicação dos modelos.

Os dados para validação foram classificados em arquivos com os dados do trimestre e em arquivos com dados mensais, visando verificar alguma oscilação sazonal mensal entre os

valores obtidos. Cabe salientar, que assim como em todos os estudos que buscam padrões no mercado financeiro, existem quatro fatores que interferem nos dados, sendo os fatores sazonais, cíclicos, de tendência e imprevisíveis.

O período amostrado passou por diversas oscilações, inclusive com a falência de mercado imobiliário norte americano, que interferiu nas negociações e cotações das bolsas de valores do mundo todo e a interferência de governos de diversos países no mercado financeiro com intuito de amenizar as perdas, através da compra e venda de dólares, por exemplo. Outros fatores como a falência de diversas instituições financeiras contam como variáveis neste ambiente. Todos estes fatos são classificados como fatores imprevisíveis, afetando qualquer estudo e ações no ambiente financeiro (ABELÉM,1994).

A validação é realizada utilizando os modelos que obtiveram nos testes a melhor classificação geral, aplicando os modelos nos dados da Bovespa de Julho a setembro de 2008. Os resultados obtidos são mostrados a seguir:

**Tabela 5.19:** Resultados de validações na Bovespa.

| Arquivo | Padrão | Trimestre | Julho | Agosto | Setembro | Média: | Técnica |
|---|---|---|---|---|---|---|---|
| Maior Ou Menor D+5 | 74,19% | 69.84 % | 76,19% | 61,90% | 71,42% | 69,84% | J48 |
| Maior Ou Menor D+2 | 75,81% | 87,30% | 85,71% | 90,48% | 85,71% | 87,30% | |
| Maior Ou Menor D+5 | 74,88% | 68,25% | 71,43% | 57,14% | 76,19% | 68,25% | MLP |
| Maior Ou Menor D+2 | 75,81% | 87,30% | 85,71% | 90,48% | 85,71% | 87,30% | |

**Fonte:** Do autor.

É possível observar que a busca de relações para o dia seguinte foi eficaz nas duas técnicas utilizadas, ou seja, com uso do arquivo TREMaiorOuMenor0D+2, de forma que ambas obtiveram na classificação média correta dos arquivos de validação 87,30%. Assim, ambas as técnicas mostram-se aptas na identificação de padrões entre os atributos do dia anterior e a variação do fechamento deste dia para o dia seguinte à data abordada.

Além disso, na classificação de dados do dia posterior, ou seja, no estudo do arquivo MaiorOuMenorD+2, nota-se que as oscilações do mercado afetam as classificações, pois no mês de agosto as duas técnicas estudadas indicaram um padrão diferente no mercado, com classificações corretas para 90,48% dos registros, contra 85,71% das classificações obtidas nos meses de julho e setembro.

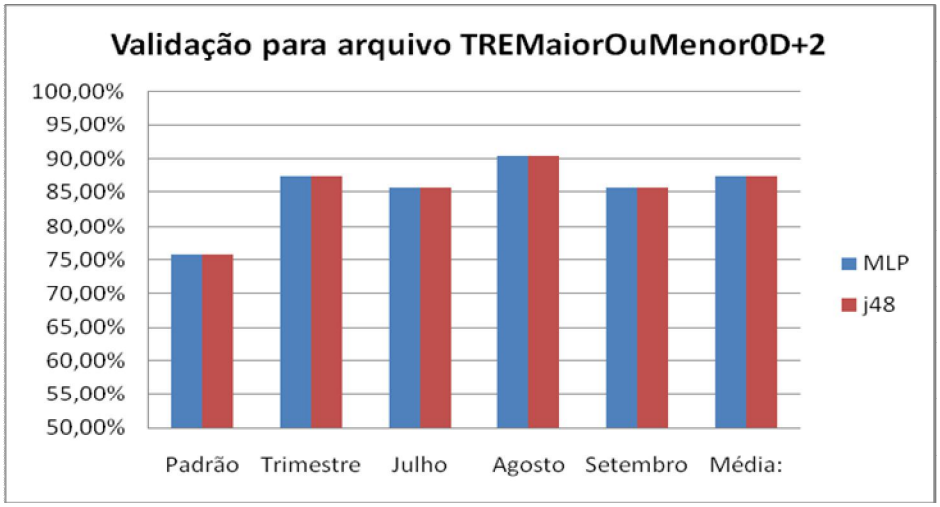


**Gráfico 5.1:** Variação para arquivo TREMaiorOuMenor0D+2
**Fonte:** Do autor.

É demonstrado que a técnica de classificação de árvores de decisão com o algoritmo J48, neste estudo de caso, é superior a técnica de redes neurais abordada para a verificação de padrões que envolvem um período maior de tempo entre os atributos relacionados, ou seja, o estudo do arquivo TREMaiorOuMenor0D+5. Foi demonstrado que o algoritmo J48 na média dos resultados de validação, apresentou 69,84% de classificações corretas contra 68,25% de classificações corretas indicadas pela técnica *multilayer perceptron.*

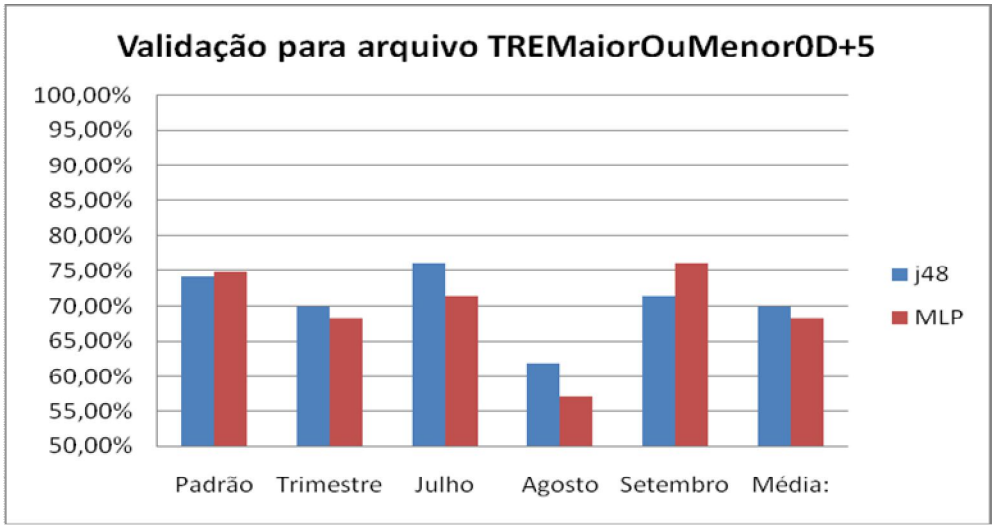


**Gráfico 5.2:**Variação para arquivo TREMaiorOuMenor0D+5
**Fonte:** Do autor.

Por fim, a validação dos algoritmos, dentro do contexto selecionado mostra que as duas técnicas são eficazes na identificação de padrões no mercado financeiro, possibilitando

assim, a exploração de tais padrões para o auxílio na tomada de decisões dos investidores que utilizam a análise técnica.

### 5.4 Exemplo Prático

Um exemplo prático é o teste dos modelos gerados pelos arquivos em um dia de movimentações na bolsa, como por exemplo, o dia 9/9/2008:

**Tabela 5.20**: Período de exemplo dos ativos abordados.

| Data | Petr4 | | | | IBOV | | Dólar | Dow Jones | |
|---|---|---|---|---|---|---|---|---|---|
| | Abertura | Volume | IFR | Fechamento | Abertura | Volume | Abertura | Abertura | Volume |
| 5/9/08 | 30,76 | 687.858.822 | 98,0392 | 31,86 | 51.404 | 4.897.338 | 1,728 | 11185,63 | 198302933 |
| 8/9/08 | 33,00 | 764.909.934 | 98,1308 | 30,26 | 51.939 | 5.148.248 | 1,725 | 11224,87 | 272995323 |
| 9/9/08 | 29,76 | 778.081.427 | 98,5816 | 28,35 | 50.710 | 5.038.642 | 1,739 | 11514,73 | 257302583 |
| 10/9/08 | 29,20 | 1.516.156.058 | 98,6301 | 28,68 | 48.439 | 6.245.922 | 1,776 | 11233,91 | 214262233 |
| 11/9/08 | 29,00 | 1.117.108.446 | 98,7261 | 31,40 | 49.628 | 5.638.889 | 1,807 | 11264,44 | 247822897 |
| 12/9/08 | 31,67 | 919.532.743 | 78,5714 | 33,00 | 51.267 | 5.263.033 | 1,791 | 11429,32 | 238882218 |
| 15/9/08 | 30,59 | 1.098.605.555 | 76,2195 | 29,80 | 52.353 | 6.570.104 | 1,831 | 11416,37 | 432965227 |
| 16/9/08 | 28,50 | 1.027.064.541 | 79,2135 | 31,30 | 48.403 | 6.466.954 | 1,822 | 10889,67 | 494755488 |
| 17/9/08 | 31,10 | 1.075.372.431 | 82,1256 | 29,80 | 49.218 | 7.457.965 | 1,811 | 11056,58 | 463197582 |

**Fonte:** Do autor.

A determinação das faixas é realizada levando-se em consideração os valores de variações do dia anterior (08/09/2008) dos atributos abertura_Petr4,Vol_Petr4,IFR,Abert_Ibov,Aber_Dolar,Abert_Dj, Vol_DJ e Fech_Petr4. Além destes atributos, é considerada a variação do fechamento da ação da Petr4 do dia, no caso 09/09/08. Logo:

Com relação ao dia 08/09/2008:

Abert_Petr4: (33,00/30,76)-1 = 7,28%, sendo classificado como Maior0.

Vol_Petr4: (764.909.934,00/687.858.822,00)-1 = 11,20%, sendo classificado como 0a50.

IFR: 98,1308, sendo classificado como 50a100.

Abert_Ibov: (51.939/51.404)-1 = 1,04%, sendo classificado como Maior0.

Vol_Ibov: (5.148.248/4.897.338)-1 = 5,12%, sendo classificado como 0a50.

Abert_Dolar: (1,725/1,728)-1= -0,17%, sendo classificado como Menor0.

Abert_DJ: (11224,87/11185,63)-1= 0,35%, sendo classificado como Maior0.

Vol_DJ: (272995323 /198302933)= 37,67%, sendo classificado como 0a50.

Fech_Petr4: (30,26/31,86)-1 = -5,02%, sendo classificado como Menor0.

O fechamento do dia 09/09/2008:

Fech_Petr4: (28,35 /30,26)-1 = -6,31%, sendo classificado como Menor0.

**Quadro: 5.5 :** Tratamento de atributos para exemplo do arquivo Tremaioroumenor0D+2
**Fonte:** Do autor.

Os atributos são organizados em um arquivo do tipo “arff” para aplicação do algoritmo, como mostrado a seguir:

```
@relation treMaiorOuMenor0-weka.filters.unsupervised.attribute.Remove-R12

@attribute Abert_Petr4_Dia_Ant {Menor0,Maior0}
@attribute VOL_Petr4_Dia_Ant {Menor-100,-100a-50,-
50a0,0a50,50a100,Maior100}
@attribute IFR_Petr4_Dia_Ant {Menor-100,-100a-50,-
50a0,0a50,50a100,Maior100}
@attribute Abert_IBOV_Dia_Ant {Menor0,Maior0}
@attribute VOL_IBOV_Dia_Ant {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute Abert_DOLAR_Dia_Ant {Menor0,Maior0}
@attribute Abert_DJI_Dia_Ant {Menor0,Maior0}
@attribute VOL_DJI_Dia_Ant {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute Fech_Dia_Ant {Menor0,Maior0}
@attribute Fech_Hoje {Menor0,Maior0}
@attribute Var_Dia_AntEPost{Menor0,Maior0}

@data
Maior0,0a50,50a100,Maior0,0a50,Menor0,Maior0,0a50,Menor0,Menor0,Menor0
```

**Quadro 5.6:** Cabeçalho de arquivo de exemplo.
**Fonte:** Do autor.

Nota-se que para aplicar os atributos selecionados no modelo, deve ser determinado um possível valor para o atributo alvo, no caso Var_Dia_AntEPost. Este atributo visa identificar a tendência do fechamento do dia posterior com relação ao dia anterior ao dia estudado. Ao aplicar o arquivo mostrado acima no modelo gerado pelo arquivo TREMaiorOuMenor0D+2,para as tecnologias de árvores de decisão e redes neurais, respectivamente, foram obtidas as seguintes respostas:

| J48 | Redes Neurais |
|---|---|
| === Summary === | === Summary === |
| Correctly Classified Instances 1 100 % | Correctly Classified Instances 1 100 % |
| Incorrectly Classified Instances 0 0 % | Incorrectly Classified Instances 0 0 % |
| Kappa statistic 1 | Kappa statistic 1 |
| Mean absolute error 0.2823 | Mean absolute error 0.4016 |
| Root mean squared error 0.2823 | Root mean squared error 0.4016 |
| Total Number of Instances 1 | Total Number of Instances 1 |

**Quadro 5.7:** Resultados De Classificações Corretas Tremaioroumenor0D+2.
**Fonte:** Do autor.

Estas respostas indicam que a variação entre o fechamento do dia posterior e o dia anterior será negativa (Menor0). Com esta informação, o investidor sabe que o valor de fechamento não ultrapassará o valor do dia anterior de R$ 30,26. Tomando como base os valores de fechamento, na melhor das situações, indicando um ganho máximo em uma operação de compra de 6,73% ((30,26/28,35)-1)*100.

Para testar a confiabilidade do modelo apresentado neste estudo, foi alterado o valor do atributo alvo Var_dia_AntEPost para Maior0, e reavaliados os modelos. O resultado indicou corretamente a classificação incorreta do arquivo, conforme mostrado a seguir:

| J48 | Redes Neurais |
|---|---|
| === Summary === | === Summary === |
| Correctly Classified Instances 0 0 % | Correctly Classified Instances 0 0 % |
| Incorrectly Classified Instances 1 100 % | Incorrectly Classified Instances 1 100 % |
| Kappa statistic 0 | Kappa statistic 0 |
| Mean absolute error 0.7177 | Mean absolute error 0.5984 |
| Root mean squared error 0.7177 | Root mean squared error 0.5984 |
| Total Number of Instances 1 | Total Number of Instances 1 |

**Quadro 5.8:** Resultados De Classificações Incorretas Tremaioroumenor0D+2.
**Fonte:** Do autor.

Visando exemplificar o modelo de classificação destacado pelos testes no arquivo TREMaiorOuMenor0D+5, foram abordados dados, com referência ao dia 10/09/2008. A determinação das faixas é realizada levando-se em consideração os valores de variações diários de dois dias anteriores (08/09/2008) dos atributos abertura_Petr4, Vol_Petr4, IFR, Abert_Ibov, Aber_Dolar, Abert_Dj, Vol_DJ e Fech_Petr4. Além destes atributos, é considerada a variação do fechamento da ação da Petr4 do dia, anterior, no caso 09/09/08 e a variação do fechamento do dia (10/09/2008) com relação a dois dias antes, no caso 08/09/2008. Este estudo visa buscar a variação o atributo fechamento da Petrobrás três dias a

frente (15/09/2008) com relação a dois dias antes ( 08/09/2008). Logo, os atributos são calculados da seguinte forma:

Com relação ao dia 08/09/2008:

Abert_Petr4: (33,00/30,76)-1 = 7,28%, sendo classificado como Maior0.

Vol_Petr4: (764909934/687858822)-1 = 11,202%, sendo classificado como 0a50.

IFR: 98,6301, sendo classificado como 50a100.

Abert_Ibov: (51.939/51404)-1 = 1,04%, sendo classificado como Maior0.

Vol_Ibov: (5.148.248,00/4.897.338,00)-1 = 5,12 %, sendo classificado como 0a50.

Abert_Dolar: (1,725/1,728)-1= -0,17%, sendo classificado como Menor0.

Abert_DJ: (11224,87/11185,63)-1= 0,35 %, sendo classificado como Maior0.

Vol_DJ: (272995323/ 198302933)= 37,67%, sendo classificado como 0a50.

Fech_Petr4: (30,26/ 31,86)-1 = 5,02%, sendo classificado como Maior0.

O fechamento do dia 09/09/2008:

Fech_Petr4: (28,35 /30,26)-1 = -6,31%, sendo classificado como Menor0.

A variação entre o fechamento do dia 08/09/2008 e o fechamento do dia 10/09/2008.

VarD-2eHoje: (28,68 /30,26)-1 = -5,22%, sendo classificado como Menor0.

**Quadro 5.9: Tratamento de atributos para exemplo do arquivo Tremaioroumenor0D+5**
**Fonte:** Do autor.

Os atributos são organizados em um arquivo para aplicação do algoritmo, como mostrado a seguir.

```
@relation treMaiorOuMenor0-weka.filters.unsupervised.attribute.Remove-R12

@attribute Abert_Petr4_Dia_Ant {Menor0,Maior0}
@attribute VOL_Petr4_Dia_Ant {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute IFR_Petr4_Dia_Ant {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute Abert_IBOV_Dia_Ant {Menor0,Maior0}
@attribute VOL_IBOV_Dia_Ant {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute Abert_DOLAR_Dia_Ant {Menor0,Maior0}
@attribute Abert_DJI_Dia_Ant {Menor0,Maior0}
@attribute VOL_DJI_Dia_Ant {Menor-100,-100a-50,-50a0,0a50,50a100,Maior100}
@attribute Fech_Dia_Ant {Menor0,Maior0}
@attribute Fech_Hoje {Menor0,Maior0}
@attribute Var_Dia_AntEPost{Menor0,Maior0}
@attribute Var_D-2eD+3{Menor0,Maior0}


@data
Maior0,0a50,50a100, Maior0 , 0a50, Menor0, Maior0,0a50,Maior0, Menor0,
Menor0,Menor0
```

**Quadro: 5.10:** Cabeçalho para exemplo de arquivo TREMaiorOuMenorD+5.
**Fonte:** Do autor.

Da mesma forma que o arquivo TREMaiorOuMenor0D+2 testado anteriormente, é necessário determinar um valor para o atributo alvo para o arquivo TREMaiorOuMenor0D+5, no caso o atributo Var_D-2eD+3. Este atributo visa identificar a tendência do fechamento de três dias à frente a data abordada, com relação ao fechamento de dois dias antes da data abordada. Os exemplos foram realizados utilizando as tecnologias de árvores de decisão e redes neurais, onde respectivamente, foram obtidos os seguintes resultados:

| J48 | Redes Neurais |
|---|---|
| === Summary === | === Summary === |
| Correctly Classified Instances 1 100 % | Correctly Classified Instances 1 100 % |
| Incorrectly Classified Instances 0 0 % | Incorrectly Classified Instances 0 0 % |
| Kappa statistic 1 | Kappa statistic 1 |
| Mean absolute error 0.3051 | Mean absolute error 0.3259 |
| Root mean squared error 0.3051 | Root mean squared error 0.3259 |
| Total Number of Instances 1 | Total Number of Instances 1 |

**Quadro 5.11: Resultados De Classificações Corretas Tremaioroumenor0D+5.**
**Fonte:** Do autor.

Estas respostas indicam que a variação entre o fechamento do dia 15/09/2008 e o fechamento do dia 08/09/2008 será negativa (Menor0). Com esta informação, o investidor sabe que o valor de fechamento de três dias a frente não ultrapassará o valor de fechamento do dia 08/10/2008 de R$ 30,26. Tomando como base o valor de fechamento do dia abordado, dia 10/09/2008 de R$ 28,68, a variação do fechamento pode trazer um ganho de no máximo 5,51% (((30,68/28,68)-1)*100). Desta forma, o investidor pode definir se efetiva ou não a operação de compra.

Se observado o período de 08/09/2008 a 17/09/2008, esta informação é de extrema importância, no dia 12/09/2008. Neste dia, o preço de fechamento do ativo da Petrobrás foi de R$33,00. Como o modelo indica para o próximo dia útil um fechamento abaixo de R$ 30,26, o investidor pode entrar “vendido” no mercado, ganhando na queda do preço do ativo. Neste exemplo, o ganho estimado mínimo seria mais de 9,0%. No dia posterior, o preço de fechamento deste ativo foi de R$ 29,80, 9,7% abaixo do preço de fechamento do dia 12, confirmando os padrões encontrados.

Para testar a confiabilidade do modelo apresentado neste estudo para este período amostrado no arquivo TREMaiorOuMenor0D+5, foi alterado o valor do atributo alvo para

Maior0, e reavaliados os modelos. O resultado indicou corretamente a classificação incorreta do arquivo, conforme mostrado a seguir:

| J48 | Redes Neurais |
|---|---|
| === Summary === | === Summary === |
| Correctly Classified Instances 0 0 % | Correctly Classified Instances 0 0 % |
| Incorrectly Classified Instances 1 100 % | Incorrectly Classified Instances 1 100 % |
| Kappa statistic 0 | Kappa statistic 0 |
| Mean absolute error 0.6949 | Mean absolute error 0.6741 |
| Root mean squared error 0.6949 | Root mean squared error 0.6741 |
| Total Number of Instances 1 | Total Number of Instances 1 |

**Quadro 5.12: Resultados De Classificações Incorretas Tremaioroumenor0D+5**
**Fonte:** Do autor.

Os dois exemplos acima observaram momentos estáveis do mercado de tendências de alta ou baixa. No exemplo a seguir, o período amostrado está em um período crítico de queda, onde o dia abordado indica uma falsa recuperação do mercado, e onde estes padrões não se aplicam.

Nesta observação o fechamento do dia abordado (12/09/2008) é superior ao dia anterior (11/09/2008) em mais de 5,0%, mas o fechamento do dia seguinte indica uma queda, com relação à data abordada de 9,7%.

Aplicando o modelo abordado neste estudo para verificar os padrões do mercado para o dia seguinte, ou seja, o padrão estudado com o arquivo TREMaorOuMenor0D+2, nota-se que ao atribuir um valor de variação negativo (Menor0) para o atributo alvo, as técnicas indicam uma classificação incorreta. Com isso, o resultado do dia posterior, conforme os padrões identificados devem ser superiores ao valor do dia anterior (11/09/2008) de R$ 31,40. Porém como verificado o fechamento do dia posterior (15/09/2008) foi de R$ 29,80, não comprovando o padrão encontrado.

| J48 | Redes Neurais |
|---|---|
| === Summary === | === Summary === |
| Correctly Classified Instances 0 0 % | Correctly Classified Instances 0 0 % |
| Incorrectly Classified Instances 1 100 % | Incorrectly Classified Instances 1 100 % |
| Kappa statistic 0 | Kappa statistic 0 |
| Mean absolute error 0.7987 | Mean absolute error 0.6741 |
| Root mean squared error 0.7987 | Root mean squared error 0.6741 |
| Total Number of Instances 1 | Total Number of Instances 1 |

**Quadro 5.13: Resultados De Classificações Incorretas Tremaioroumenor0D+5**
**Fonte:** Do autor.

Este resultado já é esperado, se observadas as estruturas das árvores geradas pelo algoritmo J48 (figura 5.4). Isto comprova que os padrões encontrados funcionam somente em períodos de tendência do mercado, sendo ineficientes em períodos de oscilações diárias.

# CONCLUSÃO

Ao término deste estudo que verificou-se que possível utilizar técnicas de *data mining* no mercado à vista brasileiro como uma ferramenta de apoio à decisão por investidores que buscam lucros em curto prazo. Entretanto, assim como em qualquer estudo neste ambiente, esta abordagem também sofre interferências de fatores cíclicos, sazonais, de tendência e imprevisíveis, que podem ter influenciado nos resultados encontrados no estudo. Fatores estes, teoricamente nulos, conforme os princípios da analise técnica.

É notado que para avaliação do valor de fechamento do dia seguinte as duas técnicas abordadas são validadas, conforme os resultados apresentados. Esta informação é valiosa para o investidor que pode aproveitar uma pequena variação do mercado e obter lucros nesta oscilação. Já para um período de alguns dias à frente, o modelo mostra queda na eficiência.

Os modelos identificados, assim como diversas técnicas e modelos existentes hoje são deficientes na identificação de padrões em um período de diversas oscilações diárias com fechamentos alternados de altas e baixas consecutivos. Assim, afirmar o uso de DM como única ferramenta de tomada de decisão é precipitado. Além disso, as técnicas gráficas e estatísticas fornecem ao investidor parâmetros já conhecidos, de fácil visualização, que competem com a tecnologia de mineração de dados.

O uso da técnica de redes neurais mostra que apesar de um funcionamento “caixa-preta”, a tecnologia é eficiente na identificação de padrões no contexto aplicado, onde o percentual de acerto de classificações corretas foi similar ao da tecnologia de árvores de decisão, de processamento claro.

A técnica de classificação de árvores de decisão, por permitir a visualização de sua lógica, pode apresentar resultados significativos mesmo com um baixo percentual de acerto geral no arquivo testado. Este fato pode trazer a geração de regras isoladas com percentuais de acertos altos, o que não acontece com a técnica de RNA, cujo resultado final deve ser expressivo. Com isso, a identificação de padrões raros, mas com alto grau de confiabilidade, são facilmente identificados com o uso da técnica de árvores de decisão. Além disso, um estudo mais aprofundado na técnica J48 pode trazer melhores resultados de classificações.

A abordagem de *data mining* neste contexto indica a possibilidade do estudo em um conjunto de ativos diferentes, que podem fazer parte do mesmo setor ou cadeia produtiva. Com isso, através do comportamento do ativo de uma empresa, por exemplo, poderá ser identificada a tendência das cotações das empresas que são seus clientes.

Durante o desenvolvimento deste estudo houve algumas dificuldades como a dificuldade em aprender o funcionamento e lógica das RNAs, o que tomou muito tempo de pesquisa; a falta de conhecimento e experiência com mineração de dados, fazendo com que o tempo de preparação dos dados fosse elevado, e quem sabe, muito acima dos 70% estimados por especialistas da área; a falta de conhecimento aprofundado com o software weka, o que possibilitaria novos rumos e melhores classificações no estudo. Além disso, permitiria a classificação geral dos arquivos diretamente no software que na última versão possibilita a exploração de arquivos.csv.

Um fator crítico neste estudo foi a falta de experiência ou a falta de um especialista na área de contexto. Apesar de contatos feitos com autores de trabalhos similares (sem resposta) um especialista da área poderia ser de grande utilidade na seleção de atributos e parametrização de valores coerentes e validação dos modelos, agilizando o trabalho de pesquisa. Além disso, poderia identificar no contexto abordado as oscilações imprevisíveis e seus impactos. Como especialista do contexto abordado foi consultado diversas vezes um investidor que atua no mercado à vista, com uso da análise técnica.

Outro fator foi a dificuldade inicial em se obter dados de fontes confiáveis e nos períodos desejados. Esta dificuldade teve impacto direto no estudo, pois como a Petrobrás é uma empresa exploradora de petróleo, certamente o preço do barril de petróleo cru tem forte

influência sobre sua cotação. Entretanto não foi possível a obtenção de dados históricos deste item.

Como trabalhos futuros podem ser citados o desenvolvimento de um software de apoio à decisão com o uso dos modelos encontrados neste estudo; o desenvolvimento de uma funcionalidade de captação de dados automática de fontes confiáveis do mercado com a possibilidade de preparação dos dados de forma automática; o desenvolvimento de um sistema especialista de monitoramento do mercado, utilizando as regras internas geradas pelo algoritmo J48 citadas e geradas por este estudo, dando ao investidor a autonomia de monitoramento do sistema; o uso da mesma metodologia de estudo na abordagem de atributos como valores de mínimas e máximas diárias dos ativos. Com isso, o investidor saberia a faixa de oscilação diária do mercado para o dia seguinte, lucrando em operações de *daytrade*; a avaliação do valor das perdas e ganhos com a adoção dos modelos apresentados neste estudo em termos reais de valores.

## BIBLIOGRAFIA


ABELÉM, Antonio J. G. **Redes Neurais Artificiais na Previsão de Séries Temporais.** Rio de Janeiro:1994.Tese (Dissertação de Mestrado) – Departamento Engenharia Elétrica, Pontifica Universidade Católica do Rio de Janeiro,1994. Disponível em: <http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/PRG_0599.SXE/8489.PDF?NrOcoSis=25094&CdLinPrg=pt>. Acesso em 20.03.2008.

AURÉLIO, M; VELLASCO, M; LOPES, C.H. **Descoberta de Conhecimento e Mineração de Dados**.1999. Disponível em <www.ica.ele.puc-rio.br/cursos/download/DM-apostila1.pdf>. Acesso em 20/03/2008.

BOVESPA[1]. **Mercado de Capitais**. São Paulo: Bovespa, 2007. Disponível em: < http://www.bovespa.com.br/Pdf/merccap.pdf>. Acesso em: 25/03/2008.

BOVESPA[2]. **Dados e Notas Bovespa**. São Paulo: Bovespa, 2007. Disponível em: < http://www.bovespa.com.br/pdf/DadosNotas.pdf >. Acesso em: 25/03/2008.

BOVESPA[3]. **Panorama da Economia Brasileira e do Mercado de Capitais**. São Paulo: Bovespa, 2007. Disponível em:< http://www.bovespa.com.br/pdf/bovpanorama.pdf>. Acesso em: 25/03/2008.

BOVESPA[4]. **Participação dos Investidores no Volume Total da Bovespa - Agosto 2007**. São Paulo: Bovespa, 2007. Disponível em: <http://www.bovespa.com.br/Mercado/RendaVariavel/PartInvest/FormConsultaMensalP.asp >. Acesso em: 12/04/2008.

BOVESPA[5]. São Paulo: BM&FBOVESPA S.A, 2008. Disponível em: < http://www.bmfbovespa.com.br/portugues/QuemSomos.asp >. Acesso em: 25/10/2008.

BRESSAN, Aureliano Angel. **Tomada de Decisão em Futuros Agropecuários com Modelos de Previsão de Séries Temporais.** São Paulo:2004. RAE-eletrônica, v. 3, n. 1, Art. 9, jan./jun. 2004. Disponível em <http://www.rae.com.br/eletronica/index.cfm?FuseAction=Artigo&ID=1914&Secao=FINANÇAS&Volume=3&Numero=1&Ano=2004>. Acessado em 25/03/2008.

BRUNI, Adriano L.; FAMÁ, Rubens. **Eficiência, Previsibilidade dos Preços e Anomalias Em Mercados de Capitais: Teoria e Evidências.** São Paulo: 1998. Caderno de pesquisas Em Administração, São Paulo. V1, n°7, 2 Trim/98. Disponível em:< http://www.ead.fea.usp.br/Cad-pesq/arquivos/c7-Art7.pdf> . Acesso em 25/04/2008

CARVALHO, Juliano V. de. **Reconhecimento de Caracteres Manuscritos Utilizando Regras de Associação**. Campina Grande: 2000. Tese (Dissertação de Mestrado) - Centro de Ciências e Tecnologia, Universidade Federal da Paraíba, 2000.

CARVALHO, Luis Alfredo Vidal de. **Data Mining : A Mineração de Dados no Marketing , Medicina, Economia, Engenharia e Administração.** Rio de Janeiro: Ciência Moderna, 2005.

CAVALCANTE,Francisco; MISUMI,Jorge Y.; RUDGE, Luiz F. **Mercado de Capitais O que é, como funciona.** Rio de Janeiro:Elsevier,2005.

COELHO, Leandro dos S.;JUNIOR, Osiris C. **Rede Neural de Base Radial Aplicada Em Previsão de Séries Temporais:Algoritmo E Aplicação**. Paraná:2000. Disponível em: < http://www.abepro.org.br/biblioteca/ENEGEP2000_E0222.PDF>. Acesso em 28.03.2008

CORRÊA, Wilson.R. PORTUGAL, M..S. **Previsão de séries de tempo na presença de mudança estrutural:redes neurais artificiais e modelos estruturais**. in.: *Economia Aplicada*, vol. 2, n° 3. 1998. Disponível em: < http://www.ufrgs.br/ppge/pcientifica/1998_03.pdf>. Acesso em 15/04/2008.

DESCHATRE, Gil Ari. **Ganhe nas bolsas com o seu micro**. Rio de Janeiro: Ciência Moderna, 1997.

DOMINGUES, Marcos A.; REZENDE, Solange O. **GART: Um Algoritmo para Generalização de Regras de Associação**. São Carlos,2005. Disponível em: < http://www.lbd.dcc.ufmg.br:8080/colecoes/wamd/2005/WAMD_1.pdf>. Acesso em 20/04/2008.

ELDER, Alexander.**Como se transformar em um operador e investidor de sucesso.** Rio de Janeiro:Campus,2005.

FERNANDES, Anita M. da Rocha. **Inteligência Artificial: Noções gerais**. Florianópolis: VisualBooks, 2005.

FERNANDES, L. G. L., PORTUGAL, M. S., NAVAUX, P. O. A. **Previsão de séries de tempo: redes neurais artificiais e modelos estruturais.** In: ENCONTRO BRASILEIRO DE ECONOMETRIA,Salvador, 1995. *Anais*... Salvador: Soc. Bras. de Econometria, 1995. Disponível em: <_www.ufrgs.br/ppge/pcientifica/1995_09.pdf >. Acesso em 26.03.2008.

FILHO,Armando M.;ISHIKAWA,Sérgio. **Mercado Financeiro e De Capitais**. São Paulo: Atlas,2000.

**FOLHA ON LINE.** INVESTIDOR pessoa física responde por 25% do giro da Bovespa**.** Folha OnLine, Mar. 2008, Disponível em :<http://www1.folha.uol.com.br/folha/dinheiro/ult91u378858.shtml>. Acesso em 20/03/2008.

FORTUNA, Eduardo. **Mercado Financeiro: Produtos e Serviços**.Rio de Janeiro:Qualitymark, 1997.

FREIMAN, José Paulo; PAMPLONA, Edson de O. **Redes Neurais Artificiais na Previsão do Valor de *Commodity* do Agronegócio**. V Encuentro Internacional de Finanzas. Santiago: 2005, Disponível em: < http://www.iepg.unifei.edu.br/edson/download/ArtFreimanChile05.pdf>. Acesso em 10/06/2008.

GIUDICI, Paolo, **Applied Data Mining: Statistic Methods for Business and Industry.** Wiley,2003.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: Um Guia Prático**. Rio de Janeiro: Campus, 2004.

GONCHOROSKI, Sidinei Pereira. **Utilização de Técnicas de KDD em um Call Center Ativo**. Novo Hamburgo:2007. Trabalho de Conclusão de Curso - Instituto de Ciências Exatas e Tecnológicas, Centro Universitário Feevale, 2007.

HAYKIN, Simon. **Redes Neurais: Princípios e prática**. Porto Alegre: Bookman, 2001.

LAZO, Juan Guillermo L. **Sistema Híbrido Genético-Neural Para Montagem e Gerenciamento de Carteiras de Ações.** Rio de Janeiro:2000.Tese (Dissertação de Mestrado) – Departamento Engenharia Elétrica, Pontifica Universidade Católica do Rio de Janeiro,2000. Disponível em: < www.maxwell.lambda.ele.puc-rio.br/cgi-bin/db2www/PRG_0651.D2W/SHOW?CdLinPrg=pt&Cont=7541:pt >. Acesso em 25/03/2008.

LUDWIG, Oswaldo; MONTGOMERY, Eduard. **Redes Neurais**. Rio de Janeiro:Ciência Moderna, 2007.

MATSURA,Eduardo. **Comprar ou Vender?**.Rio de Janeiro: Saraiva, 2007.

MELLO, Marilia T. de. **Aplicação de Redes Neurais Artificiais no Processo de Precificação de Ações.** Pelotas:2004. Instituto de Física e Matemática. Universidade Federal de Pelotas, 2004. Disponível em: <http://www.ufpel.tche.br/prg/sisbi/bibct/acervo/info/2004/mono_marilia.pdf>. Acesso em 20.06.2008.

MONGIOVI, Giuseppe. **T.E.I. Data Mining**. Notas de Aula Data Mining. Campina Grande,1998.

MONTEIRO, Manuel José Ferreira. **Classificação.** Porto:2005, Mestrado em Análise de Dados e Sistemas de Apoio à Decisão.Universidade de Porto. [online].Dez.2005, Disponível

em: < http://www.fundacaofia.com.br/proinfo/artigos/artigotelebras.pdf>. Acesso em: 26.03.2008.

OLIVEIRA Fernando L. et al. **Utilização de Algoritmos Simbólicos para a Identificação do Número de Caroços do Fruto Pequi**, In: IV Encontro de Estudantes de Informática do Estado do Tocantins**,** 2002, Palmas. Encontro de Estudantes de Informática do Tocantins – Encoinfo. Palmas, 2002.

OLIVEIRA, Ivana C. de. **Aplicação de Data Mining na Busca de um Modelo de Prevenção da Mortalidade Infantil**. Florianópolis: 2001. Tese (Dissertação de Mestrado) – Engenharia e Sistemas, Universidade Federal de Santa Catarina, 2001. Disponível em: < http://teses.eps.ufsc.br/defesa/pdf/6643.pdf>. Acesso em: 15/04/2008.

PAULOS, John Allen. **A lógica do mercado de ações**. Rio de Janeiro:Campus,2004.

PEDRO, Lucilene M; GUERREIRO, Reinaldo. **Aplicação de Árvores de Decisão na Análise Financeira.** São Paulo:2004. 1° Congresso USP Iniciação científica em Contabilidade. Disponível em:< www.congressoeac.locaweb.com.br/artigos12004/424.pdf>. Acesso em: 15/06/2008.

PORTUGAL, M. S., FERNANDES, L. G. L. **Redes neurais artificiais e previsão de séries econômicas:uma introdução**. *Nova Economia*, v.6, n.1, p.51-73, 1996. Disponível em < http://www.ufrgs.br/ppge/pcientifica/1995_01.pdf>. Acesso em 20.03.2008

PRODANOV, Cleber C. **Manual de Metodologia Científica.** 3. ed. Novo Hamburgo: Editora Feevale. 2006.

PUC- Rio. **Sistemas de Mineração de Dados**. Rio de Janeiro: 2004. Certificação Digital Nº 0210427/CA. Disponível em: <www.maxwell.lambda.ele.puc-rio.br/cgi-bin/PRG_0599.EXE/5303_3.PDF?NrOcoSis=14032&CdLinPrg=pt>. Acessado em 15/06/2008.

REZENDE, S.O. **Mineração de Dados**. XXV Congresso da Sociedade Brasileira da Computação [online]. Jul.2005, Disponível em:

<http://www.sbc.org.br/bibliotecadigital/download.php?paper=417>. Acessado em 02. Jan.2008.

RIOS, Kleyson. **FIBONACCI NO MERCADO DE AÇÕES.** 2008. Disponível em: <http://kleysonrios.blogspot.com/2008/01/sequncia-de-fibonacci-no-mercado-de-aes.html>. Acesso em 12/04/2008.

ROMÃO et al, **Extração de Regras de Associação em C&T: O Algoritmo Apriori.** Florianópolis, 1999. Disponível em: < http://www.din.uem.br/wesley/Apriori.pdf>. Acesso em 20/04/2008.

SILVA, Marcelino P. dos S. **Mineração de dados – Conceitos, Aplicações e Experimentos com Weka**. Mossoró: 2006. Universidade do Estado do Rio Grande do Norte, 2006. Disponível em: <www.sbc.org.br/bibliotecadigital/download.php?paper=35>. Acesso em: 01/04/2008.

SANTOS, Tiago. **Simulação Multiagentes Aplicado ao Mercado de Capitais**.Novo Hamburgo:2007,Instituto de Ciências Exatas e Tecnológicas, CENTRO UNIVERSITÁRIO FEEVALE, 2007.

ZANETI, L.A.; ALMEIDA, F.C. de. **Exploração do uso de redes neurais na previsão do comportamento de ativos financeiros.** Terceiro Seminário em Administração. [online]. Out.1998, Disponível em:
< http://www.fundacaofia.com.br/proinfo/artigos/artigotelebras.pdf>. Acesso em: 26.03.2008.

**Anexo A:** Testes Com algoritmo J48

| **Parâmetro:** | **teste01** | **teste02** | **teste03** | **teste04** | **teste05** | **teste06** | **teste07** | **teste08** | **teste09** | **teste10** | **teste11** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Binary Splits* | False | False | False | False | True | False | False | False | False | False | False |
| *Confidence Factor* | 0,25 | 1 | 0,5 | 0,1 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 |
| *Debug* | False | False | False | False | False | True | False | False | False | False | False |
| *MinNumObj* | 2 | 2 | 2 | 2 | 2 | 2 | 10 | 20 | 2 | 2 | 2 |
| *NumFolds* | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 10 | 20 | 20 |
| *ReducedErrorPruning* | False | False | False | False | False | False | False | False | False | False | True |
| *SaveInstanceData* | False | False | False | False | False | False | False | False | False | False | False |
| *Seed* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *SubtreeRaising* | True | True | True | True | True | True | True | True | True | True | True |
| *Unpruned* | False | False | False | False | False | False | False | False | False | False | False |
| *UseLaplace* | False | False | False | False | False | False | False | False | False | False | False |
| TreMaiorOuMenor0D+2 | 75,81% | 65,58% | 73,25% | 75,81% | 75,12% | 75,81% | 75,81% | 75,81% | 75,81% | 75,81% | 72,32% |
| TreMaiorOuMenor0D+5 | 72,32% | 65,58% | 66,51% | 73,72% | 72,56% | 72,33% | 74,19% | 73,72% | 72,33% | 72,33% | 73,26% |
| **Média** | 74,07% | 65,58% | 69,88% | 74,77% | 73,84% | 74,07% | 75,00% | 74,77% | 74,07% | 74,07% | 72,79% |

| Parâmetro: | **teste12** | **teste13** | **teste14** | **teste15** | **teste16** | **teste17** | **teste18** | **teste19** | **teste20** | **teste21** | **teste22** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Binary Splits* | False | False | False | False | False | False | False | False | False | False | False |
| *Confidence Factor* | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0.1 |
| *Debug* | False | False | False | False | False | False | False | False | False | False | False |
| *MinNumObj* | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 10 | 20 | 2 | 10 |
| *NumFolds* | 10 | 3 | 3 | 3 | 3 | 3 | 10 | 3 | 3 | 3 | 3 |
| *ReducedErrorPruning* | True | True | False | False | False | False | False | False | False | False | False |
| *SaveInstanceData* | False | False | False | False | False | False | False | False | False | False | False |
| *Seed* | 1 | 1 | 5 | 10 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *SubtreeRaising* | True | True | True | True | False | True | True | True | True | True | True |
| *Unpruned* | False | False | False | False | False | True | True | True | True | False | False |
| *UseLaplace* | False | False | False | False | False | False | False | False | False | True | False |
| TreMaiorOuMenor0D+2 | 74,19% | 73,49% | 75,81% | 75,81% | 75,81% | 66,05% | 66,05% | 71,86% | 74,18% | 75,81% | 75,81% |
| TreMaiorOuMenor0D+5 | 71,40% | 70,93% | 72,32% | 72,32% | 72,32% | 65,58% | 65,58% | 70,23% | 71,86% | 72,33% | 73,72% |
| **Média** | 72,79% | 72,21% | 74,07% | 74,07% | 74,07% | 65,81% | 65,81% | 71,05% | 73,02% | 74,07% | 74,77% |

**ANEXO B:** Regras Geradas pelo algoritmo J48.

| Regra | Acertos | Erros | Soma | Resultados | Percentual | Arquivo Origem |
|---|---|---|---|---|---|---|
| Se Hoje-2 = -2a0 E Hoje-7 = 2a0 Então Amanhã= 2a0 | 116 | 48 | 164 | 116,0/48,0 | 70,73 | TREHoje |
| Se Abert_Petr4 = 2a0 E Abert_IBOV = 0a-2 Então Fech_Dia_post = 0a5 | 102 | 41 | 143 | 102,0/41,0 | 71,32 | treinamento 5 |
| Se Abert_Petr4 = -2a-4 Então Fech_Dia_post = 0a5 | 64 | 23 | 87 | 64,0/23,0 | 73,56 | treinamento 5 |
| Se Abert_Petr4 = 0a-2 E VOL_Petr4 = 0a25 Então Fech_Dia_post = 0a5 | 64 | 23 | 87 | 64,0/23,0 | 73,56 | treinamento 5 |
| Se VarD-2 = Maior3 Então VarD+3 = Maior3 | 131 | 31 | 162 | 131,0/31,0 | 80,86 | Treinamento Dias da Semana |
| Se VarD-2 = Menor-3 Então VarD+3 = Menor-3 | 108 | 43 | 151 | 108,0/43,0 | 71,52 | Treinamento Dias da Semana |
| Se FechamentoHoje = Maior3 Então VarD+3 = Maior3 | 54 | 16 | 70 | 54,0/16,0 | 77,14 | Treinamento Dias da SemanaSem D+2 |
| Se VarD-2 = Maior3 Então VarD+3 = Maior3 | 146 | 38 | 184 | 146,0/38,0 | 79,34 | Treinamento VarD+5 |
| Se VarD-2 = Menor-3 Então VarD+3 = Menor-3 | 100 | 41 | 141 | 100,0/41,0 | 70,92 | Treinamento VarD+5 |
| Se FechamentoHoje = Maior0 Então VarD+1= Maior0 | 462 | 93 | 555 | 462,0/93,0 | 83,24 | TREMaiorOuMenor0D+2 |
| Se FechamentoHoje = Menor0 Então VarD+1 = Menor0 | 379 | 107 | 486 | 379,0/107,0 | 77,98 | TREMaiorOuMenor0D+2 |
| Se VarD-2 = Maior0 Então VarD+3 = Maior0 | 476 | 90 | 566 | 476,0/90,0 | 84,09 | TREMaiorOuMenor0D+5 |
| Se VarD-2 = Menor0 E FechamentoHoje = Menor0 Então VarD+3 = Menor0 | 272 | 83 | 355 | 272,0/83,0 | 76,61 | TREMaiorOuMenor0D+5 |
| Se FechamentoHoje = Maior3 Então VarD+1 = Maior3 | 59 | 9 | 68 | 59,0/9,0 | 86,76 | TreVarD+2 |
| Se VARD-5 = Maior7 Então VarD+5 = Maior7 | 70 | 19 | 89 | 70,0/19,0 | 78,65 | VARIAÇÃO D+10 |
| Se VARD-10 = Maior7 Então Então VarD+5= Maior7 | 191 | 39 | 230 | 191,0/39,0 | 83,04 | VARIAÇÃO D+15 |
| Se VARD-10 = Menor-6 Então VarD+5= Menor-6 | 99 | 35 | 134 | 99,0/35,0 | 73,88 | VARIAÇÃO D+15 |

# **ANEXO C:** Testes De Modelos De Redes Neurais

| **Parâmetros/Teste** | **T 01** | **T 2** | **T 3** | **T 4** | **T 5** | **T 6** | **T 7** | **T 8** | **T 9** | **T 10** | **T 11** | **T 12** | **T 13** | **T 14** | **T 15** | **T 16** | **T 17** | **T 18** | **T 19** | **Teste 20** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GUI | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| Auto Build | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Debug | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| Decay | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| Hidden Layers | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a |
| Learning Rate | 0.2 | 0.05 | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 | 0,005 | 0,05 | 0,005 | 0,005 | 0,005 | 0,005 | 0,005 | 0,02 | 0,8 | 0,05 | 0,05 | 0,05 |
| Momentum | 0.3 | 0.01 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,0001 | 0,001 | 0,001 | 1 | 0,001 | 0,001 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 |
| NominalToinaryFilter | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Normalize Attributes | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Normalize Numeric Class | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Random Seed | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Reset | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Training Time | 2000 | 2000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 10000 | 5000 | 5000 | 5000 | 5000 | 10000 | 10000 | 5000 | 5000 | 10000 | 10000 | 10000 |
| ValidationSetSize | 70 | 80 | 8 | 0 | 80 | 90 | 70 | 60 | 70 | 70 | 70 | 70 | 0 | 70 | 70 | 70 | 70 | 70 | 80 | 80 |
| Validation Threshold | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 70 |
| TREMaiorOuMenor0D+2 | 72,33% | 75,12% | 75,58% | 62,56% | 74,65% | 68,37% | 75,58% | 73,72% | 74,88% | 75,35% | 74,88% | 43,48% | 67,44% | 74,88% | 74,88% | 74,65% | 68,14% | 75,58% | 74,65% | 72,79% |
| TREMaiorOuMenor0D+5 | 70,23% | 70,47% | 74,19% | 66,98% | 70,23% | 66,97% | 72,33% | 73,26% | 72,33% | 72,39% | 72,33% | 60,44% | 64,42% | 72,33% | 72,09% | 72,56% | 66,28% | 72,32% | 70,23% | 68,60% |
| Média de Classificação : | 71,28% | 72,79% | 74,88% | 64,77% | 72,44% | 67,67% | 73,95% | 73,49% | 73,60% | 73,87% | 73,61% | 51,96% | 65,93% | 73,60% | 73,49% | 73,60% | 67,21% | 73,95% | 72,44% | 70,70% |

| **Parâmetros/Teste** | **T 21** | **T 22** | **T 23** | **T 24** | **T 25** | **T 26** | **T 27** | **T 28** | **T 29** | **T 30** | **T 31** | **T 32** | **T 33** | **T 34** | **T 35** | **T 36** | **T 37** | **T 38** | **T 39** | **T 40** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GUI | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| Auto Build | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Debug | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| Decay | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| Hidden Layers | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a | a |
| Learning Rate | 0.001 | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 | 0,05 | 0.001 | 0.001 | 0,001 | 0.00001 | 0.001 | 0,005 | 0,1 | 0,05 | 0.2 | 0.2 | 0.2 | 0,05 | 0,001 |
| Momentum | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,01 | 0,01 | 0,01 | 0,001 | 0,01 | 0,05 | 0,00001 | 0,01 | 0,01 | 0.3 | 0.3 | 0.3 | 0,1 | 0,001 |
| NominalToinaryFilter | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Normalize Attributes | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Normalize Numeric Class | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Random Seed | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Reset | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True | True |
| Training Time | 5000 | 5000 | 5000 | 5000 | 5000 | 10000 | 5000 | 5000 | 10000 | 5000 | 10000 | 10000 | 10000 | 10000 | 10000 | 5000 | 10000 | 500 | 10000 | 5000 |
| ValidationSetSize | 80 | 10 | 5 | 20 | 30 | 20 | 20 | 20 | 20 | 10 | 20 | 20 | 70 | 20 | 20 | 70 | 70 | 70 | 80 | 20 |
| Validation Threshold | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 0 | 20 |
| TREMaiorOuMenor0D+2 | 58,84% | 75,34% | 75,58% | 75,58% | 74,41% | 75,58% | 75,58% | 75,81% | 75,81% | 75,81% | 72,79% | 75,81% | 74,88% | 74,65% | 75,58% | 72,33% | 72,33% | 72,33% | 74,65% | 75,81% |
| TREMaiorOuMenor0D+5 | 73,02% | 74,18% | 72,79% | 74,16% | 73,48% | 74,19% | 73,95% | 73,95% | 73,95% | 74,88% | 60,93% | 73,95% | 72,33% | 73,49% | 73,95% | 70,23% | 70,23% | 70,23% | 70,23% | 73,95% |
| Média de Classificação : | 65,93% | 74,76% | 74,19% | 74,87% | 73,95% | 74,88% | 74,77% | 74,88% | 74,88% | 75,35% | 66,86% | 74,88% | 73,60% | 74,07% | 74,77% | 71,28% | 71,28% | 71,28% | 72,44% | 74,88% |

## **ANEXO D:** Nomes Dos Arquivos e Descrição

| Arquivo | Descrição |
|---|---|
| TREMaiorOuMenor0D+5 | Fechamento D+5 com variação positiva ou negativa, exceto atributos de volume e IFR com variação a cada 50% |
| TREMaiorOuMenor0D+2 | Fechamento D+2 com variação positiva ou negativa, exceto atributos de volume e IFR com variação a cada 50% |
| treinamento0a3 | Fechamento D+5 com variação a cada 3% de 12 a -12, exceto atributos de volume e IFR com variação a cada 50% |
| treinamento0a3DiaPost | Fechamento do dia posterior com variação a cada 3% de 12 a -12, exceto atributos de volume e IFR com variação a cada 50% |
| treinamento0a3DiaPost_Sem IFR | Fechamento do dia posterior com variação a cada 3% de 12 a -12, exceto atributos de volume com variação de 50%, sem o atributo IFR |
| treinamento0a5 | Fechamento D+5 com variação a cada 5% de 10 a -10, exceto atributos de volume e IFR com variação a cada 50% |
| treinamento0a5diaPost | Fechamento do dia posterior com variação a cada 5% de 10 a -10, exceto atributos de volume e IFR com variação a cada 50% |
| Treinamento5 | Fechamento do dia posterior com variação a cada 2% de 10 a -10, exceto atributos de volume com variação a cada 25% e sem IFR |
| Treinamento | Fechamento do dia posterior com variação a cada 2% de 10 a -10, exceto atributos de volume com variação a cada 25% e IFR a cada 10% |
| Treinamento_FaixasReclass | Fechamento D+5 com variação a cada 4% de 8 a -8, exceto atributos de volume com variação a cada 20%, sem IFR |
| Treinamento_FaixasReclassDiaPosterior | Fechamento do dia posterior com variação a cada 4% de 8 a -8, exceto atributos de volume com variação a cada 20%, sem o atributo IFR |
| Treinamento_PETR | Fechamento da Petr4 com relação a abertura (ambos com variação a cada 1% na faixa de 5 a -5) e volume de Petr4 a cada 25% de -100 a 100%. |
| Treinamento_SemIFR | Fechamento D+5 com variação a cada 2% de 10 a -10, exceto atributos de volume com variação a cada 25% sem IFR. |
| Treinamento_SemIfrDiario | Fechamento diário com variação a cada 1% de 5 a -5, exceto atributos de volume com variação a cada 25% sem IFR. |
| Treinamento_VolReclass | Fechamento D+5 com variação a cada 2% de 10 a -10, exceto atributos de volume com variação a cada 50% e sem IFR |
| TReVarD+2 | Fechamento D+2 a cada 1% de -3 a 3, exceto atributos de volume e IFR com variação a cada 50%. |
| TReVarD+5 | Fechamento D+5 a cada 1% de -3 a 3, exceto atributos de volume e IFR com variação a cada 50%. |
| TreHoje | Relaciona o fechamento dos últimos 9 dias da Petrobrás com o fechamento do dia posterior com faixas a cada 2% de 12 a -8%. |
| TreHoje+5 | Relaciona o fechamento dos últimos 10 dias da Petrobrás com o fechamento de 5 dias a frente com faixas a cada 2% de 12 a -8%. |
| TreHoje+10 | Relaciona o fechamento dos últimos 10 dias da Petrobrás com o fechamento de 5 e 10 dias a frente com faixas a cada 2% de 12 a -8%. |
| TreHoje+15 | Relaciona o fechamento dos últimos 10 dias da Petrobrás com o fechamento de 5, 10 e 15 dias a frente com faixas a cada 2% de 12 a -8%. |
| TreHoje+20 | Relaciona o fechamento dos últimos 10 dias da Petrobrás com o fechamento de 5, 10, 15 e 20 dias a frente com faixas a cada 2% de 12 a -8%. |
| Treinamento Sexta | Fechamento D+5 a cada 1% de -3 a 3, exceto atributos de volume e IFR com variação a cada 50%, avaliando apenas valores de sextas-feiras. |
| Treinamento_Quinta | Fechamento D+5 a cada 1% de -3 a 3, exceto atributos de volume e IFR com variação a cada 50%, avaliando apenas valores de quintas-feiras. |
| Treinamento_Quarta | Fechamento D+5 a cada 1% de -3 a 3, exceto atributos de volume e IFR com variação a cada 50%, avaliando apenas valores de quartas-feiras. |
| Treinamento_Terça | Fechamento D+5 a cada 1% de -3 a 3, exceto atributos de volume e IFR com variação a cada 50%, avaliando apenas valores de terças-feiras. |
| Treinamento_Segunda | Fechamento D+5 a cada 1% de -3 a 3, exceto atributos de volume e IFR com variação a cada 50%, avaliando apenas valores de segundas-feiras. |
| Treinamento_SegundaSemD+5 | Fechamento D+2 a cada 1% de -3 a 3, exceto atributos de volume e IFR com variação a cada 50%, avaliando apenas valores de segundas-feiras. |
| treinamento Tendência D+1 | Relaciona o fechamento do dia posterior com os valores de abertura e fechamento do dia atual e o fechamento dos últimos cinco dias da Petr4. |
| treinamento Tendência D+2 | Relaciona o fechamento de D+2 com os valores de abertura e fechamento do dia atual,fechamento do dia posterior e o fechamento dos últimos cinco dias da Petr4. |
| treinamento Tendência D+3 | Relaciona o fechamento de D+3 com os valores de abertura e fechamento do dia atual,fechamento do dia posterior , D+2 e o fechamento dos últimos cinco dias da Petr4. |
| treinamento Tendência D+5 | Relaciona o fechamento de D+5 com valores de abertura e fechamento do dia atual,fechamento do dia posterior , D+2, D+3 e o fechamento dos últimos cinco dias da Petr4. |
| treinamento Tendência D+10 | Relaciona o fechamento de D+10 com abertura e fechamento do dia atual,fechamento do dia posterior , D+2, D+3, D+5 e o fechamento dos últimos cinco dias da Petr4. |
| treinamento Tendência D+15 | Relaciona o fechamento de D+15 com abertura e fechamento do dia atual,fica. do dia posterior , D+2, D+3, D+5, D+10 e o fechamento dos últimos cinco dias da Petr4. |
| treinamento Variação D+1 | Com base nas variações de cinco dias anteriores busca identificar relações com o dia seguinte com faixas a cada 1% de 6 a -6%. |
| treinamento Variação D+2 | Com base nas variações de cinco dias anteriores, e dia posterior busca identificar relações com D+2 com faixas a cada 1% de 6 a -6%. |
| treinamento Variação D+3 | Com base nas variações de cinco dias anteriores, dia posterior e D+2 busca identificar relações com D+3 com faixas a cada 1% de 6 a -6%. |
| treinamento Variação D+5 | Com base nas variações de cinco dias anteriores, dia posterior, D+2 e D+3 busca identificar relações com D+5 com faixas a cada 1% de 6 a -6%. |
| treinamento Variação D+10 | Com base nas variações de cinco dias anteriores, dia posterior, D+2, D+3 e D+5 busca identificar relações com D+10 com faixas a cada 1% de 6 a -6%. |
| treinamento Variação D+15 | Com base nas variações de cinco dias anteriores, dia posterior, D+2, D+3,D+5 e D+10 busca identificar relações com D+15 com faixas a cada 1% de 6 a -6%. |