PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA DE CIÊNCIAS EXATAS E DA COMPUTAÇÃO
CURSO DE CIÊNCIA DA COMPUTAÇÃO

MINERAÇÃO DE DADOS APLICADA A PREVISÃO DE PREÇOS DE AÇÕES UTILIZANDO WEKA

WANDERSON BLEINER COELHO DE SOUZA

GOIÂNIA
2021

WANDERSON BLEINER COELHO DE SOUZA

MINERAÇÃO DE DADOS APLICADA A PREVISÃO DE PREÇOS DE AÇÕES UTILIZANDO WEKA

Trabalho de Conclusão de Curso apresentado à Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Sibelius Lellis Vieira
Banca examinadora: Prof. Me. Geraldo Valeriano Ribeiro
Prof. Me. Joriver Rodrigues Canedo

GOIÂNIA
2021

WANDERSON BLEINER COELHO DE SOUZA

MINERAÇÃO DE DADOS APLICADA A PREVISÃO DE PREÇOS DE AÇÕES UTILIZANDO WEKA

Trabalho de Conclusão de Curso aprovado em sua forma parcial pela Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, para obtenção do título de Bacharel em Ciência da Computação, em _____/_____/_____.

_____

Orientador: Prof. Dr. Sibelius Lellis Vieira

_____

Prof. Me. Geraldo Valeriano Ribeiro

_____

Prof. Me. Joriver Rodrigues Canedo

GOIÂNIA

2021

RESUMO

Este trabalho visa aplicar técnicas de mineração de dados no âmbito da bolsa de valores, visando predizer preços futuros de ações, auxiliando investidores do mercado acionário em suas tomadas de decisões. Uma pesquisa bibliográfica sobre o mercado acionário e mineração de dado, foi realizada com objetivo de entender e conceituar o mercado de ações e técnicas de mineração de dados. Foi feito um levantamento dos dados históricos da ação PETR4, em que se extraiu dados referentes ao ano de 2019 desses dados pelo site Yahoo *Finance.* A seleção, limpeza e estruturação desses dados foram realizadas com auxílio de um programa e também manualmente. Com auxílio do *software* WEKA, foi realizada a mineração de dados e seus resultados apresentados e analisados, indicando sua aplicação para prever se a ação tem comportamento de aumento ou diminuição de preço dentro de um intervalo definido de dias.

Palavras-chave: bolsa de valores, mercado acionário, ciência de dados; mineração de dados, descoberta de conhecimento em base de dados, redes neurais, árvore de decisão.

ABSTRACT

This work aims to apply data mining techniques in the field of the stock market, aiming to predict future stock prices, assisting stock market investors in their decision making. A literature search on the stock market and data mining was conducted with the aim of understanding and conceptualizing the stock market and data mining techniques. A survey was made of the historical data of the PETR4 share, in which data relating to the year 2019 were extracted by the Yahoo Finance site. The selection, cleaning and structuring of this data was done with the help of a program and also manually. With the help of WEKA software, data mining was performed and its results presented and analyzed, indicating its application to predict whether the stock has price increase or decrease behavior within a defined interval of days.

Keywords: stock market, data science; data mining, neural networks, decision tree.

LISTA DE ABREVIATURAS

ARFF	*Attribute-Relation File Format*
B3	Brasil, Bolsa, Balcão
BM&FBovespa	Bolsa de Mercadorias e Futuros de São Paulo
CSV	*Comma-separated Values*
CVM	Comissão de Valores Mobiliários
DCBD	Descoberta de Conhecimento em Base de Dados
IBOVESPA	Índice da Bolsa de Valores de São Paulo
IRF	Índice de Força Relativa
KDD	*Knowledge Dicovery in Databases*
LSE	*London Stock Exchange*
MME	Média Móvel Exponencial
MMS	Média Móvel Simples
NASDAQ	National Association of Securities Dealers Automated Quotations
NYSE	*New York Stock Exchange*
SLP	*Single Layer Perceptron*
SVM	*Support Vector Machine*
WEKA	*Waikato Environment for Knowledge Analysis*

LISTA DE ILUSTRAÇÕES

FIGURAS

QUADROS E TABELAS

SUMÁRIO

1 INTRODUÇÃO

1.1 Contextualização

De acordo com a Comissão de Valores Mobiliários (2019), o mercado de capitais possui um grande papel na economia de um país, pois é fonte de captação de recursos de investidores através da bolsa de valores, o que possibilita financiar atividades de empresas, estimulando a economia. Segundo d'Ávila (2021), o número de investidores tem crescido no Brasil, houve um crescimento de 92% no número de investidores pessoas físicas na bolsa de valores brasileira, aumentando em 1,5 milhão de investidores em 2020.

O lucro dos investidores são frutos da valorização de ativos, e para maximizar os lucros, investidores tem buscado cada vez mais por ferramentas que os auxiliem na tomada de decisão, em relação a compra ou venda determinada ação. Prever o mercado acionário é extremamente complexo, pois existem diversas variáveis que podem dificultar, como questões políticas, expectativas dos investidores e até mesmo desastres naturais. Mesmo assim é possível predizer preços futuros com um bom grau de confiança (MARANGONI, 2010).

Existem vários estudos que utilizam inteligência artificial para predizer tendências ou preços de ações. Segundo Castro e Ferrari (2016), a mineração de dados utiliza diversas técnicas de inteligência artificial, com objetivo de descobrir conhecimento que podem estar escondidos em uma grande quantidade de dados. A aplicação da mineração de dados cabe em diversas áreas de conhecimento, inclusive na predição de dados históricos, como é o caso da previsão de preços de ações.

Redes neurais artificiais é umas das técnicas utilizadas na mineração de dados, que possui uma grande capacidade de generalização, por conta de sua habilidade de aprender. Essa capacidade, em conjunto com sua estrutura paralelamente distribuída, faz com que uma rede neural seja capaz de produzir resultados impressionantes para problemas considerados complexos (HAYKIN, 2001).

Outra técnica muito utilizada na mineração de dados é a árvore de decisão, que trabalha com classificações de dados. Sua capacidade de ser fácil de se explicar e de fácil compreensão, a tornam umas das melhores tarefas de se trabalhar na mineração de dados, além disso, sua construção é feita de maneira bem simples, o que faz com que ela seja bem concisa (CASTRO; FERRARI, 2016).

Esse trabalho, tem como proposta, explorar o funcionamento da bolsa de valores e da mineração de dados, e aplicar o processo de mineração de dados nos preços históricos de fechamento da ação da Petrobrás (PETR4), a fim de prever o preço futuro da ação e então sugerir que seja feita a compra ou venda da ação. E por fim, analisar seus resultados.

1.2 Justificativa

Nesse contexto, a necessidade dos investidores da bolsa de valores de obter lucros com a compra e venda de ações, se mostra importante o uso de ferramentas que os auxiliem na tomada de decisão, o que justifica a investigação da aplicabilidade da mineração de dados a preços de ações e a análise dos resultados obtidos a partir de experimentos com suas técnicas.

1.3 Objetivos

1.3.1 Objetivo geral

Aplicar e analisar técnicas de mineração de dados na predição do preço de fechamento da ação da Petrobrás (PETR4) usando o software WEKA, com vistas a determinar se a ação deve ser comprada ou vendida dentro de um intervalo definido de dias.

1.3.2 Objetivos específicos

Para se atingir o objetivo geral, propõem-se os seguintes objetivos específicos:

- Conceituar o mercado de ações;
- Conceituar a mineração de dados;
- Aplicar o processo de descoberta de conhecimento em base de dados;
- Analisar resultados obtidos a partir da aplicação da mineração de dados utilizando o WEKA;
- Analisar aplicabilidade da mineração de dados no âmbito do mercado acionário.

1.4 Estrutura do trabalho

Esse estudo apresenta-se estruturado em 5 (cinco) capítulos, sendo o capítulo 1 (um) referente a esta introdução. No segundo capítulo é apresentado toda pesquisa bibliográfica do trabalho, na qual são definidos conceitos do âmbito do mercado financeiro e do processo de descoberta de conhecimento em base de dados. O capítulo 3 (três) é composto pelos materiais e métodos utilizados nesse trabalho de pesquisa. No capítulo 4 (quatro), são apresentados os resultados gerados pela pesquisa e na sequência, esses resultados são discutidos. Por fim o capítulo 5 (cinco), apresenta a conclusão da pesquisa.

2 REFERENCIAL TEÓRICO

2.1 Mercado financeiro

Esta seção tem como objetivo apresentar conceitos sobre bolsa de valores, o propósito e funcionamento do mercado financeiro, abordando o conceito de ações, questões relativas ao preço das ações, movimentação de compra e venda, a teoria de Dow e a análise técnica.

2.1.1 Bolsa de valores

De acordo com Bratti (2009), a bolsa de valores é o local no qual as empresas negociam suas ações em capital aberto. A imagem evocada quando se trata de bolsa de valores, é a de um ambiente em que pessoas agitadas com um telefone grande na mão realizam compras e vendas de ativos. Porém, com o grande avanço das tecnologias tem se transformado, pois essas operações já podem ser feitas a partir de um celular, embora essas movimentações só possam ser feitas através de uma corretora que intermedia a operação. A bolsa de valores também tem a fama de ser o local em que pessoas fazem ou perdem fortunas do dia para a noite.

Cada país possui suas próprias bolsas de valores. Por exemplo, o Estados Unidos possui a *National Association of Securities Dealers Automated Quotations* (NASDAQ) e a *New York Stock Exchange* (NYSE), Londres tem a *London Stock Exchange* (LSE), que estão entre as maiores bolsas de valores mundiais. A única Bolsa de Valores brasileira, B3 (sigla que significa Brasil, Bolsa, Balcão), surgiu da fusão da Bolsa de Mercadorias e Futuros de São Paulo (BM&FBovespa) com a Central de Custódia e de Liquidação Financeira de Títulos (Cetip). A bolsa brasileira possui mais de 350 empresas na sua lista, e são executadas milhares de movimentações diárias de compra e venda dessas ações realizadas (ELIAS, 2020).

Segundo a Comissão de Valores Mobiliários (2019), no Brasil o responsável por fiscalizar o mercado de capitais é a própria Comissão de Valores Mobiliários (CVM). A CVM foi criada em 1976 com objetivo de fazer com que os investidores se sintam protegidos para aplicar seus investimentos, assegurando que o os mercados de bolsa e balcão sejam eficientes e regulares, coibindo manipulações no sentido de criar demandas artificiais de preços de valores mobiliários, além de garantir ao público acesso a informações em relação aos valores mobiliários emitidos.

2.1.2 Ações – conceito e tipos

Quando uma ação é negociada, por exemplo, sendo vendida a determinada pessoa, esta pessoa está adquirindo uma pequena fração do capital social de uma empresa, e assim o acionista (pessoa que comprou a ação) está sujeito aos lucros e prejuízos que a empresa obtiver. Então pode-se dizer que o acionista se torna sócio dessa empresa (FOGAÇA, 2015). Do ponto de vista contábil/econômico, a empresa tem um patrimônio total que é o seu ativo, uma dívida total que é o seu passivo e um patrimônio líquido, que é a diferença entre o ativo e o passivo. No patrimônio líquido está o capital da empresa, e a ação é parte deste capital para empresas de sociedade aberta. A figura 2.1 ilustra a relação entre a ação e patrimônio da empresa.

Figura 2.1 - Relação ação e patrimônio da empresa.

Fonte: Elaborado pelo autor.

Cada empresa pode definir se suas ações serão emitidas com ou sem valor nominal, de acordo com seu estatuto. Quando as ações são emitidas com valor nominal, todas as ações possuirão o mesmo valor e não será possível emitir novas ações com valor diferente. Ações emitidas sem valor nominal tem preço de emissão definidos pelos sócios fundadores da companhia (ASSAF, 2018).

Existem dois principais tipos de ações no mercado brasileiro: ações ordinárias e ações preferenciais.

O acionista ordinário possui o poder de influenciar em decisões da empresa, pois uma das características é o direito ao voto. Esses acionistas podem decidir sobre destinações de lucros e investimentos, elegem a diretoria da sociedade entre outros assuntos. Os detentores de ações preferenciais não possuem direito ao voto, por consequência, não participam de deliberações na empresa. No entanto desfrutam de

alguns privilégios, tais como, prioridade no recebimento dos dividendos, garantia de um valor mínimo fixo de dividendo e em caso de liquidação da sociedade tem preferência no reembolso do capital (ASSAF, 2018).

2.1.3 Preço de ações

Segundo Giacomel (2016), existem diversas formas de se analisar o desempenho de uma ação, sendo a mais comum examinando seu preço final, quando as negociações se encerram ao final do dia. No entanto, apenas essa métrica não é suficiente para saber como a ação se comportou ao longo do dia, pois diversas operações de compra e venda acontecem desde o momento da abertura do pregão até seu fechamento, fazendo com que seu preço varie para cima e para baixo. Outras métricas podem ser utilizadas para se avaliar de maneira mais profunda o preço de uma ação, e as mais usadas pelos investidores são as seguintes:

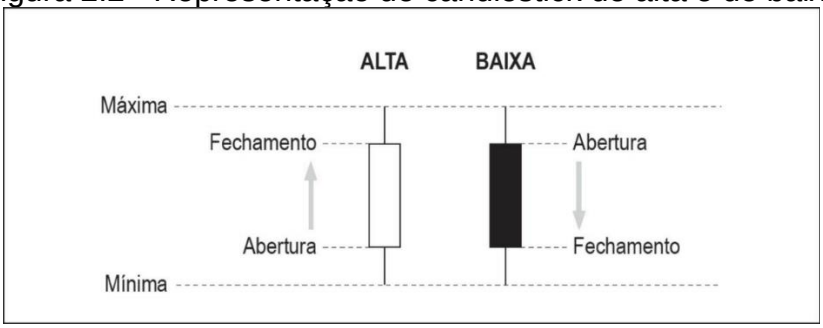
- Fechamento – é o valor atingido pela ação no momento do fechamento do pregão;
- Abertura – primeiro preço de uma ação, na abertura de um dia de negociações, nem sempre igual ao valor do fechamento do dia anterior;
- Máximo – trata-se da maior cotação do dia em meio as variações durante o pregão;
- Mínimo – menor cotação do dia;
- Volume – somam-se todas as operações de compra e venda realizadas durante o dia para chegar a esse valor.

Analisando algumas dessas métricas em conjunto é possível chegar em certas características que podem ajudar na tomada de decisão. Por exemplo, o valor do fechamento sendo muito menor que o valor máximo no dia indica uma queda ao longo do dia, e assim observa-se uma tendência de que a ação continue caindo no dia seguinte, e isso se aplica inversamente, caso o valor de fechamento seja muito maior que o valor mínimo do dia. O volume mostra-se uma ótima ferramenta para avaliar o comportamento do mercado. Por exemplo, se o volume de certa ação começa a cair após um intervalo de alta, significa que há uma diminuição de interesse do mercado nessa ação, o que leva a queda dos preços do ativo.

Uma maneira de se representar os preços de fechamento, abertura, mínimo e máximo de uma ação é usando uma representação gráfica chamada *candlestick*, que

recebe esse nome pelo formato parecido com uma vela que apresenta. O *candlestick* pode ser utilizado para retratar o comportamento de determinada ação pelo espaço de tempo que desejado, por exemplo, 10 minutos em um dia, períodos de 20 dias ou uma semana (DEBASTIANI, 2007). A figura 2.2 representa o *candlestick* de alta e de baixa.

Figura 2.2 - Representação de candlestick de alta e de baixa



Fonte: Debastiani (2007)

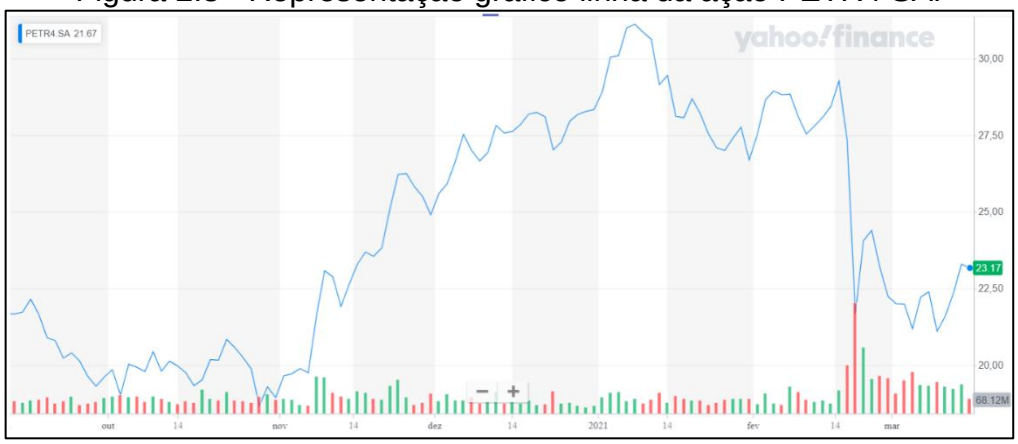
2.1.4 Tipos de gráficos

Gráficos são excelentes aliados para auxiliar na tomada de decisão em investimentos. A forma de apresentar e a quantidade de dados podem variar de acordo com cada tipo de gráfico. Dentre vários tipos, esses são alguns dos mais utilizados:

- Gráfico de linha;
- Gráfico de barras;
- Gráfico *candlestick*.

O mais simples é o gráfico de linha, e os analistas costumam utilizar desse gráfico para analisar a variação do fechamento em relação ao tempo, visto que muitos investidores acreditam que esse é o dado mais importante do dia, podendo ser usado para avaliar outros parâmetros. Esse tipo de gráfico se caracteriza pela ligação por uma linha do preço de um dia ao preço do próximo dia, trazendo uma visualização fácil da variação dos preços, sendo a forma mais eficiente de se analisar uma ação (LEMOS, 2016). A figura 2.3 representa a variação do preço do fechamento da ação PETR4 ao longo de seis meses, utilizando o gráfico linha.

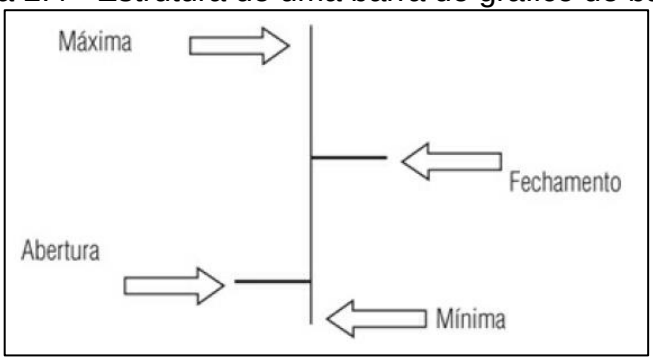
Figura 2.3 - Representação gráfico linha da ação PETR4 SA.



Fonte: Yahoo Finance.

No gráfico de barra, são utilizados os dados de abertura, fechamento, máximo e mínimo. Para representar um dia, em um gráfico diário, é utilizado uma barra na vertical, mostrando os dados do pregão no dia. Os extremos da barra representam máximo e mínimo do dia, na haste à esquerda tem-se a abertura e à direita o fechamento, como é representado na figura 2.4.

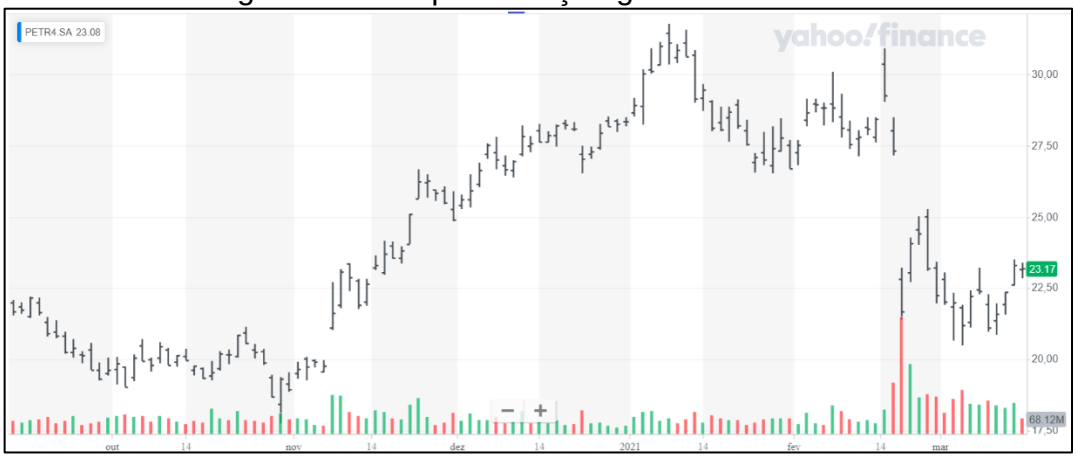
Figura 2.4 - Estrutura de uma barra do gráfico de barras.



Fonte: Lemos (2016).

Como no gráfico linha, este gráfico, além de diário, pode ser utilizado para fazer análises *intraday* (períodos de um, cinco, quinze, trinta e sessenta minutos entre um dia e outro), semanais, quinzenais, mensais (LEMOS, 2016). A figura 2.5 representa a variação dos preços da ação PETR4 ao longo seis meses, utilizando o gráfico barra.

Figura 2.5 – Representação gráfico de barras.



Fonte: Yahoo Finance

De acordo com Debastiani (2007), o *candlestick* é um dos tipos gráficos mais antigos. Os japoneses já utilizavam esse modelo no século XVII para analisar a Bolsa de Arroz em Osaka. Nos Estados Unidos chegou por volta de 1980, importado por um operador de ações em Nova York chamado Steve Nison.

Assim como no gráfico de barras, são utilizados os dados de fechamento, abertura, máximo e mínimo do ativo e sua construção é baseada em representar um dia de negociações por meio de uma “vela” (área formada entre os preços de abertura e fechamento) na vertical, os dados de máxima e mínima são representados por uma linha que passa sobre o corpo da vela, chamados “sombras”, sendo a extremidade superior o preço máximo e a inferior o mínimo como ilustrado na figura 2.2.

Um *candlestick* de alta é quando o fechamento é maior que a abertura, ou seja, significa que o ativo obteve uma valorização durante o dia, sendo representado pelo corpo do *candle* branco ou vazio, e em caso contrário, quando uma ação tem uma desvalorização ao longo do dia, se dá o nome de *candlestick* de baixa e o seu corpo é representado pela cor preta como mostra a figura 2.2. Algumas representações mais modernas podem trocar essas cores, e como na figura 2.6 que demonstra a variação da ação PETR4 ao longo de seis meses, em que um *candle* de alta é representado pela cor verde e um *candle* de baixa está em vermelho.

Figura 2.6 - Representação gráfico *candlestick*

Fonte: Yahoo Finance

2.1.5 Análise gráfica – teoria de Dow

A análise técnica tem como base a chamada teoria de Dow. Charles Dow, um dos fundadores do Dow Jones *Financial News Service*, usava seus editoriais no *The Wall Street Journal* para escrever sobre os princípios básicos do que viria a se tornar a teoria de Dow. Apesar disso, Dow não escreveu sobre sua teoria. William Hamilton, sucessor dos editoriais de Dow após a sua morte, concluiu e organizou os princípios da teoria de Dow (LEMOS, 2016).

Os índices criados por Dow, com propósito de medir a movimentação do mercado, eram formados pelas principais empresas dos setores industrial e ferroviário da época, que são o Dow-Jones Ferroviário e Dow-Jones Industrial que eram formados por grandes empresas de cada setor.

Segundo Lemos (2016), a teoria de Dow é baseada em princípios. O primeiro princípio diz que o mercado tem três tendências, ou seja, a movimentação do mercado é dividida em três tendências, a primária, a secundária e a terciária. A tendência primária chega a durar mais de um ano, e sua movimentação pode gerar grandes altas ou baixas nos preços. A tendência secundária acontece dentro da primária, alterando a tendência primária por um certo tempo, e dura entre três semanas e três meses, corrigindo o movimento da tendência primária entre um e dois terços. Na tendência terciária tem uma duração menor que as anteriores, tendo fim em no máximo três meses, são pequenas oscilações que podem ser “controladas” por grupos que possuem um poder financeiro maior.

O segundo princípio afirma que o volume deve acompanhar a tendência. A variação de preços deve ser seguida pelo volume de negociações de ações. Quando houver uma tendência de alta nos preços, o volume de negociações deve aumentar, e diminuir caso haja uma desvalorização. Em uma tendência de baixa, a valorização do ativo deve causar uma diminuição no volume de negociações, e um aumento quando o ativo estiver desvalorizado (FOGAÇA, 2015).

No terceiro princípio alega-se que as tendências primárias de alta possuem três fases. Na fase de acumulação, os investidores mais experientes e qualificados começam a comprar, pois percebem que o mercado está com preços baixos. Na subida sensível, as empresas têm um resultado melhor causando um aumento regular no preço das ações, e investidores com mais sensibilidade percebem o momento, o que faz com que o volume de negociações aumente nas altas e diminua nas quedas. Na fase denominada estouro, grande parte dos investidores percebe a tendência de alta e ganham confiança para comprar, fazendo com que os preços subam de forma acentuada (ROQUE, 2009).

Conforme Lemos (2016) afirma no quarto princípio, as tendências primárias de baixa possuem três fases. A fase de distribuição acontece ao fim de uma fase de estouro. Investidores experientes começam a vender suas ações, fazendo com que os preços comecem a cair. Na fase do pânico, há uma diminuição drástica na quantidade de compradores, e começa um processo de venda dentre os investidores que pressentem que algo pode estar errado, fazendo com que os preços tenham quedas acentuadas e o volume de negociações aumente. Na última fase, chamada baixa lenta, os preços estão muito baixos, e faz com que investidores que não venderam anteriormente se sintam desencorajados de se desfazer de suas ações.

Conforme Roque (2009), o quinto princípio diz que os índices descontam tudo. As atitudes de investidores, tanto inexperientes quanto investidores mais bem informados (com mais informações e previsões mais precisas), são refletidas pelos índices. O mercado se adapta aos acontecimentos que refletem diretamente nos preços.

No sexto princípio, de acordo com Fogaça (2015), tem-se que as duas médias devem se confirmar. Esse princípio afirma que as médias devem seguir o mesmo caminho. Um indício de mudança na tendência de uma ação pode ser a ruptura da confirmação da direção de um dos índices.

O sétimo princípio indica que o mercado pode se desenvolver em linha. A teoria de Dow diz que uma movimentação lateral pode acontecer com certa regularidade. Apesar da possibilidade de acontecer na fase de acumulação ou distribuição, a direção da linha ocorre normalmente no sentido da tendência primária (LEMOS, 2016).

O oitavo princípio afirma que as médias devem ser calculadas com preços de fechamento. Os preços de máxima, mínima e abertura, na teoria de Dow não são utilizadas, pois há um consenso entre investidores de que o valor do fechamento representa a tendência das negociações ao longo do dia (LEMOS, 2016).

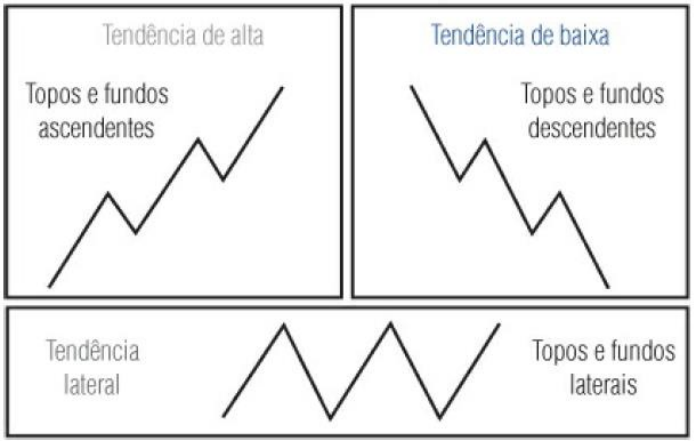
Finalmente o nono princípio diz que, a tendência está valendo até que haja sinais de reversão. Esse princípio diz que não se deve acreditar totalmente em uma reversão de tendência sem que se confirme o fim de uma tendência. Existem críticas acerca de que investidores demoram a perceber essa confirmação por parte do mercado (LEMOS, 2016).

2.1.6 Tendências

Segundo Lemos (2016), a direção em que o mercado se movimenta é chamada de tendência dos preços. As tendências podem ser classificadas em três tipos, tendência de alta, de baixa e tendência lateral, também conhecida como zona congestão. A tendência de alta se caracteriza por ter topos e fundos em ascendência, de forma que os compradores pagam valores mais altos, fazendo com que a tendência se mantenha. A tendência de baixa se caracteriza por ter topos e fundos em descendência, com os vendedores vendendo com preços mais baixos, mantendo a tendência. Já em uma tendência lateral, os níveis dos novos topos e fundos ficam praticamente iguais.

O investidor, na intenção de obter lucro, tenta identificar o momento em que uma tendência se inicia, e de posse dessa informação, ele poderá decidir o melhor momento para realizar uma operação de compra ou venda. A figura 2.7 representa as três direções das tendências.

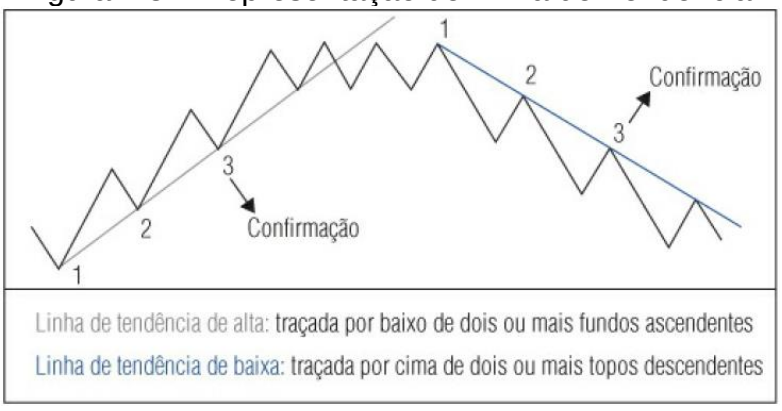
Figura 2.7 - Representação das direções das tendências.



Fonte: Lemos (2016).

Uma das formas mais eficientes e fáceis para confirmar tendências se chama linha de tendência. Uma linha projetada de baixo para cima ligando pelo menos dois fundos, é conhecida como linha de tendência de alta. Uma linha projetada de baixo para cima ligando pelo menos dois topos, representa uma linha de tendência de baixa. A confirmação da tendência vem de um terceiro ponto na linha de tendência na mesma direção, quanto mais pontos, mais forte se torna essa tendência, implicando em um impacto maior quando essa linha for quebrada (LEMOS, 2016). A figura 2.8 ilustra as linhas de tendências de alta e de baixa.

Figura 2.8 – Representação de Linha de Tendência.



Fonte: Lemos (2016).

Para considerar que uma linha de tendência está seguindo na direção principal, além do terceiro toque na linha, é necessário observar alguns pontos. Os pontos de referências devem estar a uma certa distância, para que se tenha certeza de que são

duas movimentações distintas. Uma tendência se torna insuportável se a tendência se desenvolver rapidamente, fazendo com que ela fique muito inclinada. A linha pode ser traçada por diversas vezes até que se ajuste ao movimento dos preços. Com base nessas observações, é possível afirmar que uma tendência só vai mudar de direção caso uma força maior na direção contrária mude os preços, o que pode iniciar uma nova tendência no sentido contrário. (LEMOS, 2016).

2.1.7 Indicadores técnicos

Giacomel (2016) mostra que indicadores são usados para entender comportamentos e tendências que muitas vezes não são percebidas apenas analisando os preços. Um indicador é calculado por uma fórmula que engloba os preços de máximo, mínimo, abertura e fechamento e em alguns indicadores também são incluídos o volume de uma ação.

Um indicador pode ser usado como ferramenta de alerta, para que o investidor perceba a movimentação de uma ação e a análise, pois em um cenário de queda, pode ser um sinal de quebra de suporte. Além disso pode servir como auxílio de a confirmação de outras ferramentas de análise gráfica, como *candlestick*, fazendo com que se preveja os preços futuros de uma ação (LEMOS, 2016).

Existem diversos indicadores, sendo que alguns dos mais usados são baseados em cálculos de valores passados das ações (histórico).

O mais simples é conhecido como Média Móvel Simples (MMS), que consiste apenas no cálculo aritmético dos preços diários em relação à um determinado espaço de tempo, geralmente são usados os preços de fechamento. As médias móveis são utilizadas para suavizar dados, diminuindo ruídos que podem atrapalhar numa análise de preços de uma ação, tornando mais fácil a identificação de uma tendência (GIACOMEL, 2016). A equação 2.1 mostra a fórmula para calcular a MMS, onde *n* representa a quantidade de dias e o d_n o preço do dia *n*.

$$MMS = \frac{(d_1+d_2+\cdots+d_{n-1}+d_n)}{n} \tag{2.1}$$

A Média Móvel Exponencial (MME) é semelhante à MMS, mas diferentemente da mesma, já a MME trabalha com pesos, dando uma importância maior aos dias mais

recentes, designando pesos maiores para estes. Dessa forma, é possível perceber a mudança de direção da tendência mais rapidamente que na MMS. Apesar da eficiência, a MME pode dar sinais falsos de mudança na direção, o que não quer dizer que seja melhor que a outra, pois é necessário analisar qual indicador usar para o caso analisado (LEMOS, 2016).

Para se calcular uma MME é necessário seguir três passos. Primeiramente é necessário calcular uma MMS, conforme a equação 2.1, a MMS é usada como ponto de partida no primeiro cálculo, sendo considerada a MME do período anterior. Após isso, uma fórmula é usada para calcular a ponderação, como demonstrado na equação 2.2. Só então é calculada a MME, como apresentado na equação 2.3 (LEMOS, 2016).

$$P = \frac{2}{t+1} \tag{2.2}$$

$$MME = (F - MMS\{dia\ anterior\}) * P + MME\{dia\ anterior\} \tag{2.3}$$

Onde:

P = Fator de Ponderação

MME = Média Móvel Exponencial

MMS = Média Móvel Simples

F = Fechamento do dia

t = período

Segundo Wilder (1978 apud GIACOMEL, 2016), o Índice de Força Relativa (IRF), se caracteriza por calcular o quão forte uma tendência de uma ação está. Diferentemente dos outros índices acima citados, o IRF possui uma escala que varia entre 0 e 100, sendo que quando oscila abaixo de 30 pontos, tem-se um ativo sobrecomprado, com preços muito baixos, que logo vão começar a subir, e quando oscilando acima de 70, pode-se dizer que é o ativo está sobrevendido, com preços altos, e os preços consequentemente, começarão a cair. O cálculo é dado pela fórmula 2.4, onde U representa a média dos preços em que os dias foram de alta e D a média dos preços em que os dias foram de baixa.

$$IRF = 100 - \left(\frac{100}{1+\frac{U}{D}}\right) \quad (2.4)$$

A importância do uso dos indicadores é grande para tomada de decisão, e saber usá-los não depende apenas de sua aplicação, pois é necessário saber qual o objetivo de cada um. Existem diversas técnicas que podem combinar mais de um indicador, podendo ser uma tarefa difícil para um investidor realizar.

2.2 Descoberta de conhecimento em base de dados

Nesta seção são apresentados conceitos de descoberta de conhecimento em base de dados (DCBD), discutir as fases que compõem a DCBD, além de conceitos de mineração de dados e suas técnicas que foram utilizadas no trabalho.

A frase “dados são o novo petróleo”, que segundo Schmelzer (2020), foi dita originalmente pelo matemático londrino Clive Humbly em 2006, revela o quão importantes são os dados, mas como o próprio Humbly afirma, se não for tratado e refinado, não terão utilidade. Com a grande quantidade de dados existentes, é praticamente impossível de serem tratados manualmente de maneira a extrair todo seu potencial, pois existem padrões que estão implícitos, impossíveis de serem detectados em técnicas tradicionais. Dessa forma, faz-se necessário o uso de ferramentas computacionais para que se consiga um bom resultado neste processo.

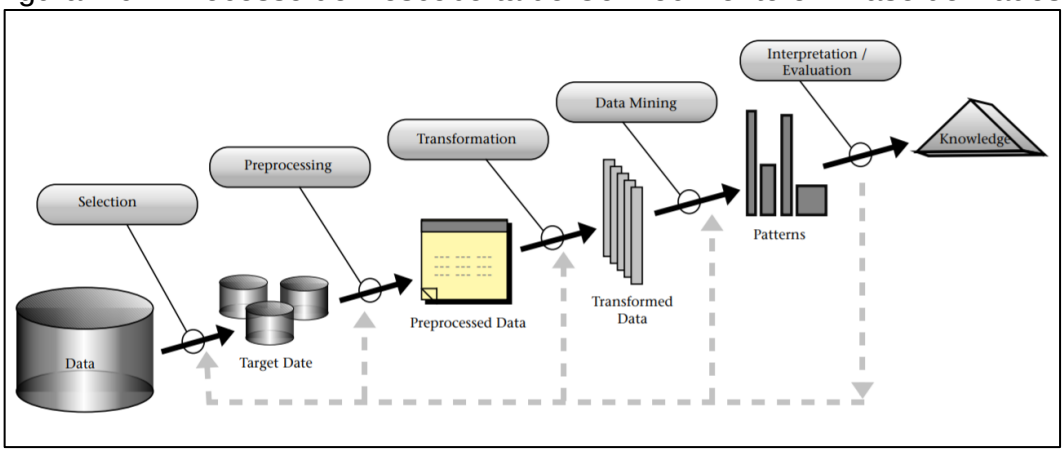
Knowledge Dicovery in Databases (KDD), ou Descoberta de Conhecimento em Bases de Dados (DCBD) é um processo que visa descobrir padrões em dados que podem estar “escondidos”, ocultos, e que podem se transformar em conhecimentos úteis para tomadas de decisões. Muitas vezes o processo DCBD é confundido com mineração de dados. No entanto, a mineração de dados é um dos passos que compõem o processo de DCBD (SILVA.; PERES; BOSCARIOLI, 2016).

2.2.1 Fases do DCBD

Segundo Soares Junior e Quintella (2005), o KDD é composto por etapas que extraem conhecimento e informações de forma não-trivial de uma base de dados, efetuando relação com suas características. São cinco etapas que compõem três

grupos principais, sendo que o primeiro grupo trata do pré-processamento, que é constituído pelas etapas de seleção dos dados, limpeza dos dados e transformação dos dados. No segundo grupo acontece a mineração de dados e no grupo do pós-processamento acontece a interpretação. A figura 2.9 mostra as cinco etapas do processo de KDD.

Figura 2.9 – Processo de Descoberta de Conhecimento em Base de Dados.



Fonte: Fayyad et al. (2016).

Na primeira etapa é necessário conhecer os dados e selecionar adequadamente quais os dados são importantes para a mineração de dados. Os dados coletados podem vir de diversas fontes, tais como, planilhas, *data warehouse*, ou até mesmo podem ter sido digitados por uma pessoa, além disso, podem estar estruturados ou não, (MOURA, 2019).

De acordo Silva (2008), a etapa de limpeza dos dados, consiste na escolha e uso de técnicas para reduzir ruídos, que podem ser desde inspeção humana, até o uso de interpolação, agrupamento ou regressão. O uso de estratégias para corrigir campos sem dados ou nulos, como a regressão e inferência podem ser utilizados nessa etapa para que se garanta a qualidade dos dados, além do uso de correção manual.

Já na terceira etapa são realizados a integração dos dados, visto que os dados podem ser de fontes diferentes, além da modificação e formatação dos dados que devem ficar adequados para a mineração de dado. A integração visa minimizar redundâncias de atributos, identificar e resolver valores conflitante em relação a diferenciação na escala ou codificação. Para transformar os dados de maneira apropriada, podem ser usadas a generalização, que transforma dados em conceitos

mais abrangentes, como faixa etária de idade, ou a normalização que cria uma escala especifica para os dados, fazendo com que os dados variem somente nessa faixa de valores, a agregação, gerando dados totalizados em relação a um determinado atributo, por exemplo, volume semanal, quinzenal ou mensal de movimentações de uma ação, e a construção de novos campos a partir de informações existentes (SILVA, 2008).

A quarta etapa é responsável pela mineração de dados, na qual são definidas e aplicadas as técnicas e algoritmos de mineração de dados que verificam a hipótese e extraem padrões de forma autônoma a partir dos dados definidos na etapa três. Nessa etapa os modelos podem ser aplicados diversas vezes, ou até mesmo refeitos, dependendo do objetivo a ser alcançado (CASTRO; FERRARI, 2016).

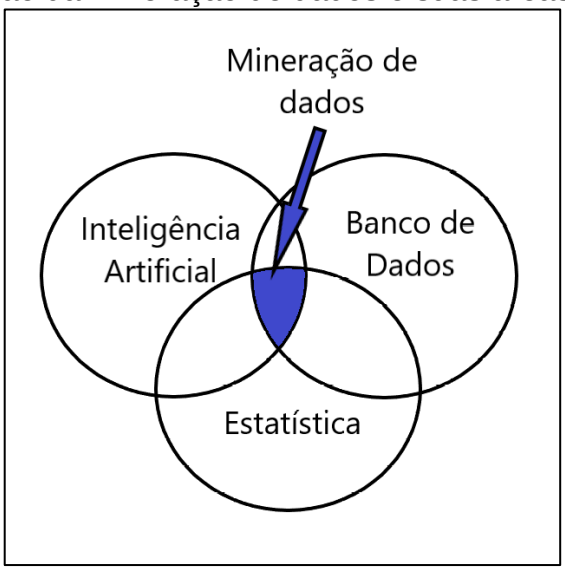
Enfim, na quinta etapa, os resultados da fase anterior são interpretados e o desempenho avaliado. A avaliação pode ser feita por meio de medidas estatísticas geradas pelos resultados. Existe a possibilidade de se retornar a qualquer uma das etapas anteriores, caso necessário.

2.2.2 Mineração de dados

Segundo Castro e Ferrari (2016), o termo mineração de dados tem como referência à extração de pedras preciosas em uma mina, no qual são usadas ferramentas adequadas para extrair essas pedras preciosas com precisão. É como se o conhecimento fosse extraído, através de algoritmos adequados, de uma base de dados. Em outras palavras, a mineração de dados é o uso de algoritmos computacionais em base de dados com objetivo de descobrir padrões úteis para tomada de decisão. O conhecimento extraído nesse processo é utilizado para tomadas de decisão, permitindo agregação de valor ao objetivo da decisão.

A mineração de dados envolve diversas áreas de conhecimentos, sendo que as principais a de banco de dados, para manipular os dados dentro das bases de dados, estatística, usada para avaliar e validar os resultados, e a inteligência artificial, sendo muito usado técnicas de aprendizado de máquina para realização da descoberta de padrões (SILVA, 2008). A figura 2.10 mostra a relação da mineração de dados com suas áreas de conhecimentos.

Figura 2.10– Relação da mineração de dados e suas áreas de conhecimento.



Fonte: Elaborado pelo autor

2.2.3 Tarefas e técnicas de mineração de dados

De acordo com Castro e Ferrari (2016), as tarefas de mineração de dados são divididas em duas categorias principais. As tarefas preditivas visam predizer valores futuros ou desconhecidos através de dados já conhecidos. Já nas tarefas descritivas, o objetivo é encontrar padrões nos dados, ou seja, descobrir características intrínsecas, de modo que seja possível a interpretação dessas características. As tarefas de classificação e regressão são categorizadas como preditivas, já as tarefas de agrupamento, modelagem de dependências, sumarização e detecção de desvios, são tarefas descritivas. Neste tópico são apresentados algumas das principais tarefas de mineração de dados.

A tarefa de classificação, consiste em analisar um novo registro e determinar uma classe com base em um histórico de registros já classificados. Por exemplo, em banco com histórico de seus clientes, pode-se utilizar esses dados para saber se um novo cliente, com certas características, irá deixar de pagar ou não um empréstimo. Essa tarefa é realizada em duas etapas, a primeira na qual é realizada o treinamento, o modelo é gerado através de um conjunto de dados já classificados. Em seguida, um novo conjunto de dados, que não foram usados na etapa anterior, é utilizado para realizar os testes, assim é possível saber se a capacidade do modelo de responder

corretamente aos dados é eficiente, essa tarefa é considera do tipo aprendizagem supervisionada (CASTRO E FERRARI, 2016).

De maneira semelhante, a regressão também é uma tarefa que segue o modelo de aprendizagem supervisionada, separando dados para o treinamento e testes. Porém, a maneira que seus resultados são avaliados é diferente, pois a regressão, por ser uma tarefa que visa predizer um valor contínuo, não observa a quantidade de acertos e erros como na classificação, e sim o cálculo da distância entre a saída esperada e a saída estimada, tendo como resultado a precisão da predição. (CASTRO E FERRARI, 2016).

A tarefa de agrupamento, como o próprio nome sugere, visa agrupar objetos em *clusters,* de acordo com relações existentes entre eles. São realizadas buscas por similaridades e diferenças entre os objetos analisados, e a distância de similaridade e diferença entre os objetos é usada para determinar em qual grupo se encaixa, ou seja, objetos similares tem distância de similaridade menor, então são agrupados em um mesmo *cluster*. Apesar de parecer com a classificação, no agrupamento não existe classes previamente determinadas, pois simplesmente, os objetos são agrupados de acordo suas semelhanças (SILVA; PERES; BOSCARIOLI, 2016).

Na associação, o objetivo é encontrar relações entre atributos que ocorrem em base de dados transacionais. A frequência de ocorrência de atributos em transações significa que existem uma relação forte entre esses atributos. Um exemplo clássico é o do carrinho de supermercado, em que pessoas que compram leite, também compram pão, indicando que então existe uma relação entre esses dois produtos (SILVA; PERES; BOSCARIOLI, 2016).

Existem diversas técnicas de mineração de dados, e cada técnica é usada de acordo com o objetivo da tarefa, dependendo do tipo de informação que se pretende obter. No Quadro 2.1 são apresentadas as relações entre algumas das principais técnicas e tarefas de mineração de dados.

Quadro 2.1 - Relação entre técnicas e tarefas de mineração de dados.

	Tarefas	**Técnicas**	**Algoritmos**
Análise preditiva	Classificação	Árvore de decisão; Análise bayesiana; análise de vizinhança; redes Neurais	J-48, algoritmo C4.5, classificadores bayesianos, KNN, SVM (*Support Vector Machine)*

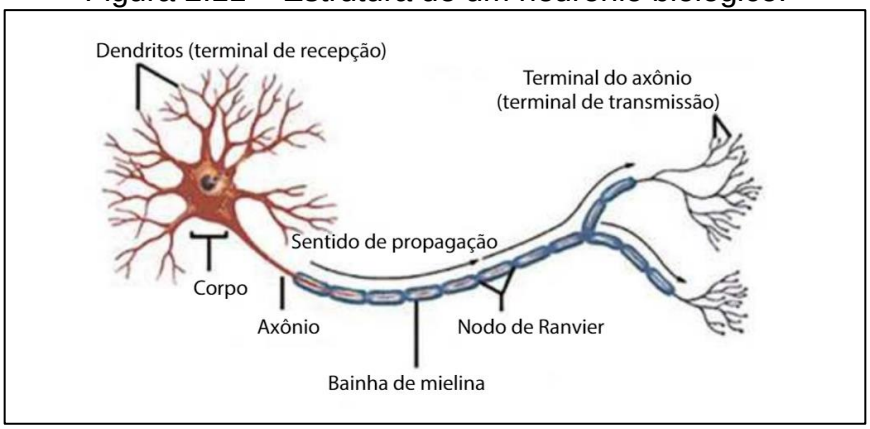
	Regressão	Regressão linear; redes neurais.	*Backpropagation; Multilayer Percetpron;*
Análise descritiva	Agrupamento	Método particionamento; modelagem de regras.	Algoritmos k-médias; *fuzzy*
	Associação	Mineração de regras de associação;	Apriori; Algoritmo FP-Growth

Fonte: Elaborado pelo autor

2.2.4 Redes neurais artificiais

O comportamento dos neurônios no cérebro humano é responsável pela inspiração da criação de modelos matemáticos chamados de redes neurais artificiais. A estrutura do neurônio biológico é composta por três partes, o corpo celular ou soma, os dendritos e o axônio, como mostra a figura 2.11. O axônio é uma fibra, que dependendo da função do neurônio, pode chegar em um metro de comprimento, em seu extremo, possui terminais, que transmitem o sinal processado pelo corpo celular para outros neurônios através dos dendritos, ou para outras partes corpo, os dendritos são responsáveis por receber os sinais de outros neurônios. As conexões formadas entre os dendritos e os terminais axônios de diferentes neurônios são chamadas de sinapses. (RUSSELL; NORVIG, 2013).

Figura 2.11 – Estrutura de um neurônio biológico.

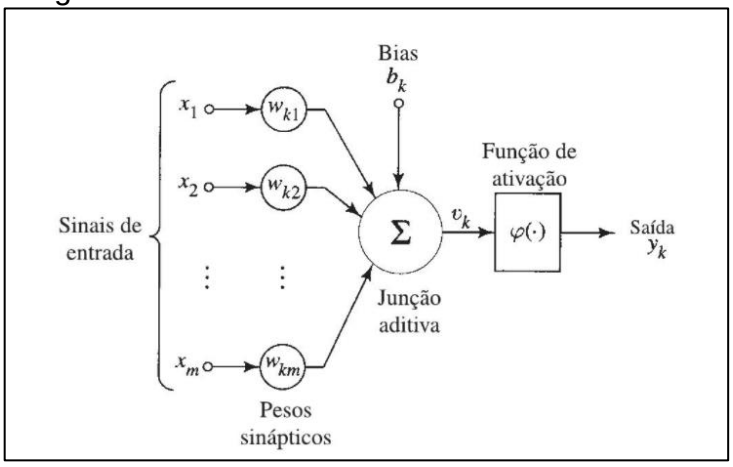


Fonte: Silva et al. (2019).

A forma mais comum de se representar uma rede neural artificial graficamente é usando grafos, em que uma aresta representa uma ligação sináptica e um nó representa um neurônio. Um neurônio pode possuir várias entradas, e para cada entrada x de um neurônio, um peso sináptico w é aplicado, sendo que os valores de

todas as entradas somados pela função soma ∑, e então uma função de ativação φ é aplicada ao resultado, o que gera uma saída, como no modelo de neurônio da figura 2.12 mostra. Essa saída pode ter como destino a entrada de um próximo neurônio quanto pode ser o valor final da rede neural (GIACOMEL, 2016).

Figura 2.12 – Estrutura de um neurônio artificial



Fonte: Haykin (2001).

A figura 2.12 mostra um modelo de neurônio que inclui a aplicação de uma constante chamada bias de maneira externa, e o objetivo do bias é fazer com que a função de ativação receba um valor mais bem adaptado aos dados, o que pode fazer com que o valor diminua, caso o bias seja negativo, ou aumente, se for positivo (HAYKIN, 2001).

A fórmula matemática da figura 2.12 é descrita pelas equações 2.5 e 2.6, no qual, em um neurônio k, u_k representa saída do combinador linear, w_{kj} representa o peso sináptico da entrada j, x_j é a entrada j e por fim saída final é representada pelo y_k.

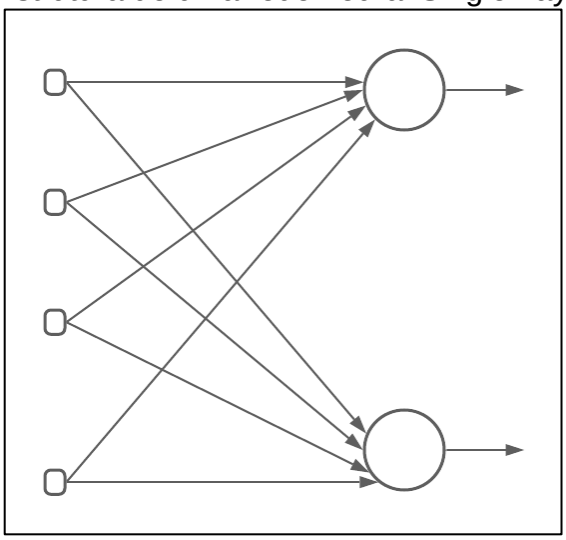
$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{2.5}$$

$$y_k = \varphi(u_k + b_k) \tag{2.6}$$

Segundo Silva et al (2019), existem diversas estruturas de redes neurais artificiais, mas duas delas a *Single Layer Percetron* e a *Multilayer Perceptron,* são consideradas as principais estruturas. A *Single Layer Perceptron* (SLP) é a mais simples, e consiste em neurônios paralelamente organizados em uma única camada,

possuindo apenas uma saída, mas podem receber n entradas, como a figura 2.13 ilustra. Dessa forma a SLP consegue resolver problemas de classificação, podendo, por exemplo, ter valores binários, 0 ou 1, como saída.

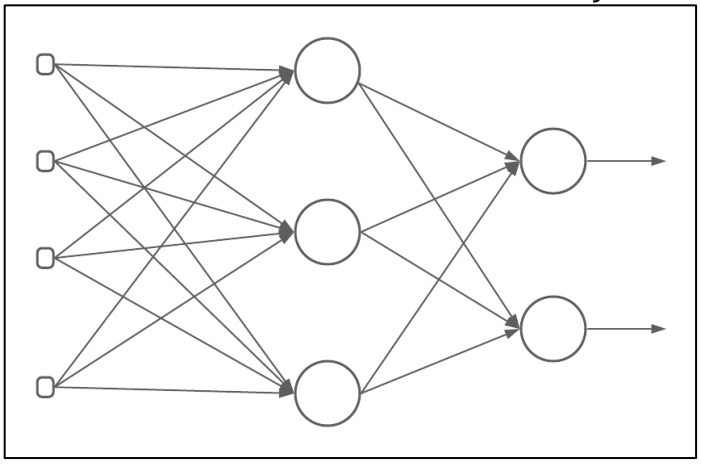
Figura 2.13 – Estrutura de uma rede neural Single Layer Perceptron.



Fonte: Elaborado pelo autor.

Já a *Multilayer Perceptron* (MLP), ou rede neural multicamadas, é um tipo de rede neural mais complexa, que pode possuir mais de uma camada de neurônio, as chamadas camadas ocultas, e os neurônios que estão na camada oculta tem como valor de entrada os resultados da camada anterior através das sinapses. As camadas ocultas conseguem aumentar o poder de processamento da rede neural, e tem objetivo de extrair resultados mais expressivos (Haykin, 2001).

Figura 2.14 - Estrutura de uma rede neural *Multilayer Perceptron.*



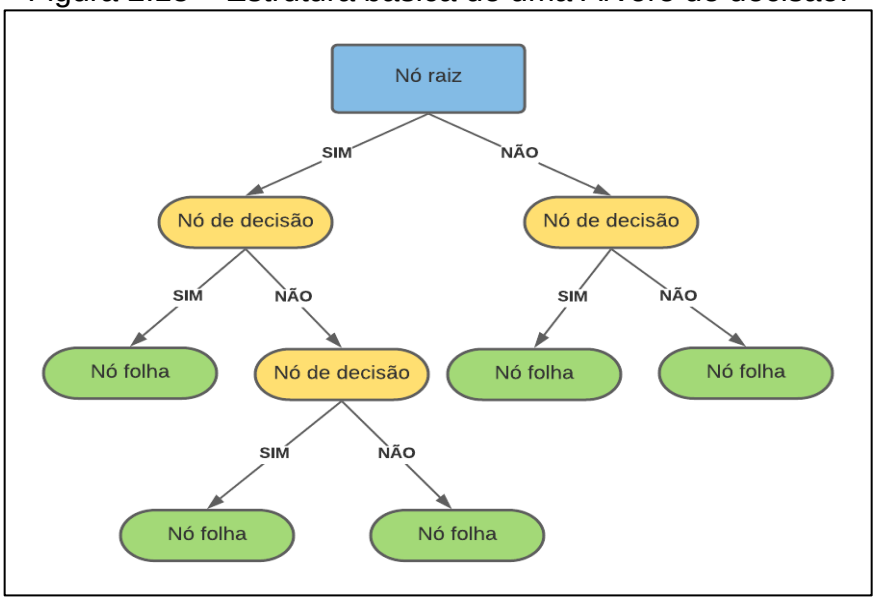
Fonte: Elaborado pelo autor.

Na figura 2.14 é adicionado uma camada oculta, com três neurônios em relação a figura 2.13, fazendo com que se torne uma rede 4-3-2, pois são 4 valores de entrada, 3 neurônios ocultos e 2 na camada de saída.

2.2.5 Árvore de decisão

Segundo Russell e Norvig (2013), uma árvore de decisão recebe como entrada um conjunto de atributos de um objeto não classificado, que podem ser contínuos ou discretos, e devolve a classificação do objeto. Uma árvore de decisão é composta por nós internos, nós folhas, ramificações e um nó raiz. O nó raiz é o nó mais alto da árvore, cada nó interno é responsável por um teste realizado em um dos atributos do objeto, os ramos de um nó são as possíveis respostas para os testes feitos no nó. A classificação final de uma árvore de decisão é responsabilidade dos nós folha, que estão ao final da árvore. A figura 2.15 ilustra a estrutura básica de uma árvore de decisão.

Figura 2.15 – Estrutura básica de uma Árvore de decisão.



Fonte: Elaborado pelo autor.

De acordo com Castro e Ferrari (2016), o processo de construir uma árvore de decisão com objetivo de classificar um objeto sem classe definida com base nos

valores do objeto, é chamado de indução de árvores de decisão. O processo de indução de uma árvore de decisão se dá de maneira recursiva:

- Um atributo é colocado na raiz da árvore e são feitas ramificações de acordo com as possibilidades dos valores, dividindo a base de dados em subclasses;
- Para cada ramo gerado, se repete o processo de maneira recursiva, só devem ser usados objetos que chegam até o ramo;
- O sinal que está no final da árvore é quando todos os objetos são classificados na mesma categoria.

Para saber qual o atributo ideal para a divisão é necessário medir a pureza dos nós, ou seja, o quão homogêneo um nó é em relação as classes do objeto. Ao se medir a pureza de todos os nós é definido como será a expansão dos nós, pois os nós com filhos mais puros são escolhidos para a expansão. A entropia é a medida que define a pureza e calcula a variabilidade das classes que pertencem ao conjunto de atributos da base de dados. (CASTRO; FERRARI ,2016).

O cálculo da entropia é definido pela equação 2.6, onde X_{tr} é um conjunto de dados de treinamento e P_{ck} refere-se à probabilidade de ocorrer a classe c_k em X_{tr}.

$$E(X_{tr}) = -\sum_{ck=1}^{k} P_{ck} \log_2(P_{ck}) \tag{2.6}$$

De acordo com Silva, Peres e Boscarioli (2016), a entropia está ligada a desordem se a entropia for alta, logo a desordem também é alta, funcionando da mesma forma quando a entropia for baixa. Então pode-se dizer que, se a entropia for alta, a dificuldade para determinar a quais classes pertencem os dados de uma base de dado será maior, demandando mais esforço para organizá-los.

Quando um nó é usado para particionar o conjunto de dados, haverá a mesma quantidade de partições que as possibilidades de classes que esse nó poderá assumir. Se todas as partições tiverem entropia igual a 0, significa que são partições puras, ou seja, cada partição terá conjuntos de dados classificados em apenas uma classe. Havendo entropia maior que 0 em alguma partição, indica que existe mais de uma classe para a partição. Então é necessário a análise da chamada informação necessária, que é o quanto de esforço é preciso para chegar em partições puras, partindo das parições atuais (SILVA; PERES; BOSCARIOLI, 2016).

O cálculo da informação necessária é dado pela fórmula 2.7, em que $E(X_{tr_i})$ é a entropia do conjunto de dados de treinamento da partição i, IN_A é a informação necessária do atributo A no conjunto de dados de treinamento total e v é o total de classes possíveis para o atributo A.

$$IN_A(X_{tr}) = \sum_{i=1}^{v} \frac{|X_{tr_i}|}{Xtr} E(X_{tr_i}) \quad (2.7)$$

E com base no conceito de informação necessária é estabelecido o conceito de ganho de informação, que é uma medida que define o quanto se ganha em colocar um determinado atributo como nó para particionar a árvore em um determinado ponto da árvore. O atributo selecionado deve ser o com maior ganho de informação (SILVA; PERES; BOSCARIOLI, 2016).

$$G(A) = E(X_{tr}) - IN_A(X_{tr}) \quad (2.8)$$

O cálculo do ganho de informação é dado pela equação 2.8 que é a subtração da entropia do conjunto de dados de treinamento $E(X_{tr})$ pela informação necessária $IN_A(X_{tr})$ do atributo A em relação ao conjunto de treinamento.

Com modelo de classificação construído, é necessário avaliar sua eficácia. Existem diversas formas para isso, sendo uma das mais conhecida a acurácia, ou taxa de classificações corretas, em que simplesmente é calculado o percentual de classificações corretas e incorretas.

Outro modelo de avaliação muito utilizado é a matriz de confusão, que é uma ferramenta usada na análise de classes, em que é possível verificar, por exemplo, se existe uma classe mais difícil de se tratar. Esse modelo compreende uma matriz com dimensões $n \times n$, em que n representa o número de classes existentes no problema. Nas linhas estão relacionadas às classificações esperadas e nas colunas as predições feitas pelos modelos (SILVA; PERES; BOSCARIOLI, 2016).

Em um problema que existem duas classes, positivo e negativo, a célula em que é indexada na linha positivo e na coluna positivo representa o número de classificações realizadas como positivo e de fatos devem ser categorizadas na classe positivo, ou seja, são verdadeiros positivos. Da mesma forma, o número de classificações que são categorizadas como negativo e realmente deveriam ser da

classe negativo são contadas na célula correspondente a linha correspondente ao negativo, coluna negativo. Nesse caso, pode-se dizer que são verdadeiros negativos, como apresentado no quadro 2.2. Igualmente acontece com os que deveriam estar classificados em uma categoria, mas foram classificados em outra categoria, e nesse caso são chamados falsos negativos e falsos positivos (SILVA; PERES; BOSCARIOLI, 2016).

Quadro 2.2 - Matriz de confusão para problema binário.

	Positivo	**Negativo**
Positivo	Verdadeiros positivos	Falsos negativos
Negativo	Falsos positivos	Verdadeiros negativos

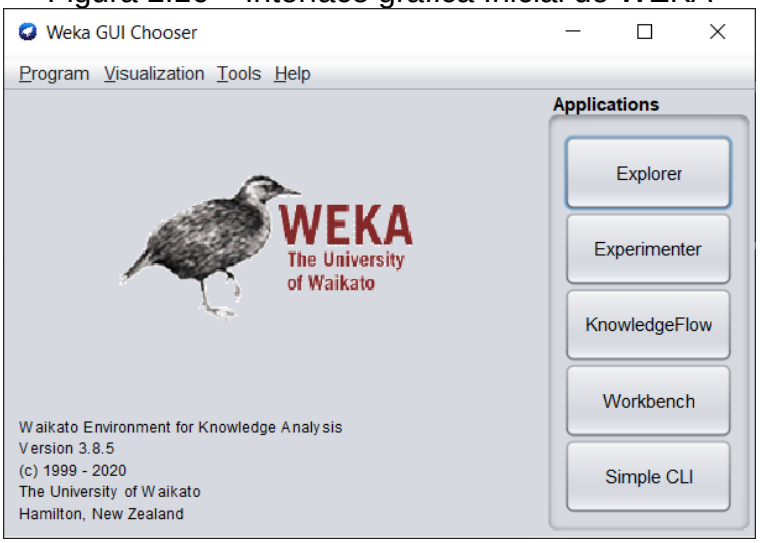
Fonte: Elaborado pelo autor.

2.3.6 Weka

Segundo Silva (2008*), Waikato Environment for Knowledge Analysis* (WEKA) é um *software* de código aberto com vários algoritmos e ferramentas de pré-processamento e mineração de dados utilizados no processo KDD. A ferramenta que tem sua origem na Nova Zelândia, mais precisamente na Universidade de Waikato, possui uma interface gráfica simples, de fácil uso, além de fornecer relatórios e histogramas dos dados.

Em sua tela inicial, o software fornece algumas opções de aplicações, *Explorer, Experimenter, KnowlwdgeFlow, Workbench e Simple CLI,* como apresentado na figura 2.16. Como o próprio nome sugere, a opção *Explorer* é uma ferramenta para explorar os dados, na qual é possível trabalhar com registros de maneira simplificada, para isso são usadas diversas opções de tratamento de dados e mineração de dados fornecidos pelo WEKA. Na aplicação *Experimenter* é possível realizar experimentos e testes estatísticos envolvendo os sistemas de aprendizagem. A função *KnowlwdgeFlow* possui aplicações parecidas com a do *Explorer,* no entanto, funciona no sistema clica e arrasta. O *Workbench* é como uma mesa de trabalho com todas as opções anteriores de aplicações em um único lugar. E por fim, a opção *Simple CLI,* diferente das outras opções, traz uma interface com linhas de comando para se trabalhar (AGOSTINI, 2017).

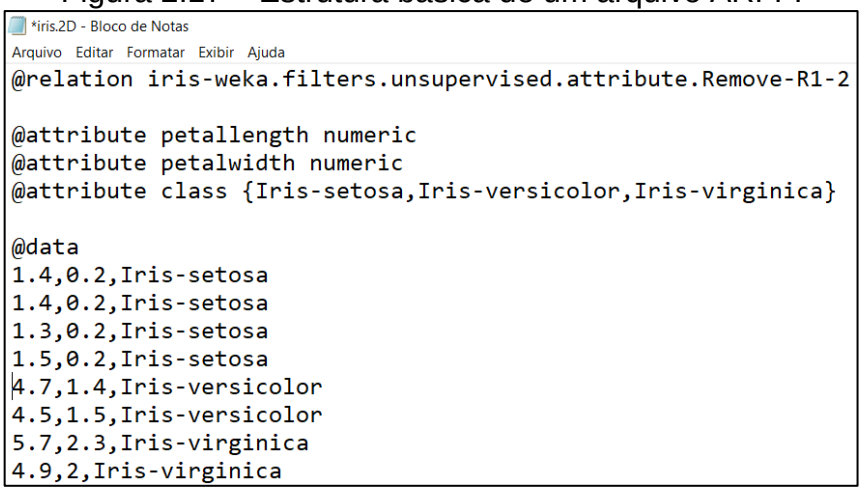
Figura 2.16 – Interface gráfica Inicial do WEKA



Fonte: Elaborado pelo autor.

Para que o WEKA reconheça os registros e seja possível manipular os dados, é necessário que os dados estejam em um arquivo no formato *Attribute-Relation File Format* (ARFF). O arquivo ARFF é um arquivo de texto com estrutura de lista, em que os registros descritos comungam dos mesmos atributos. A estrutura básica de um arquivo ARFF é formada por duas seções, conforme pode ser visto na figura 2.17, em que na primeira seção são descritas informações do cabeçalho, na qual são declarados os atributos, que podem ser do tipo *NUMERIC, DATE* ou *STRING,* e na segunda parte do arquivo se localizam os dados, que são estruturados de acordo com as declarações da primeira seção e são separados por vírgula (AGOSTINI, 2017).

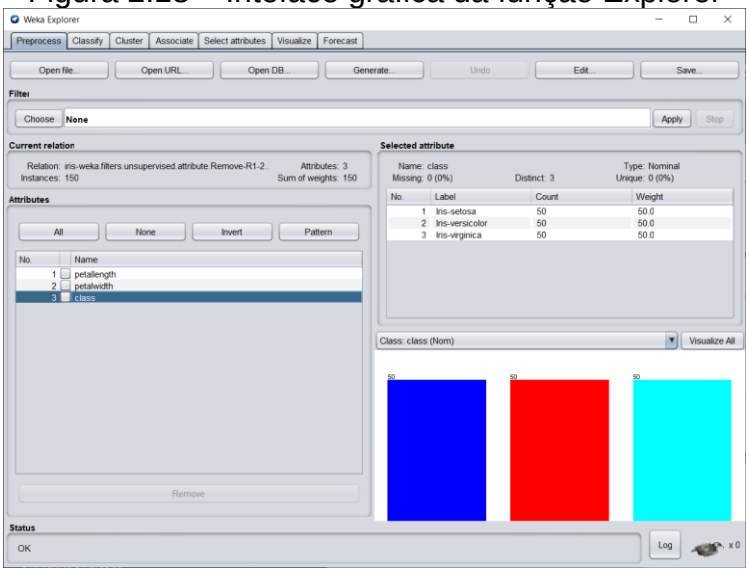
Figura 2.17 – Estrutura básica de um arquivo ARFF.



Fonte: Elaborado pelo autor.

Então, de posse de um arquivo ARFF devidamente configurado, é possível visualizar e manipular os dados, além de realizar testes com WEKA. A figura 2.18 mostra como é a visualização dos dados fornecidos por um arquivo ARFF. A função *Explorer* fornece diversas tarefas para trabalhar com mineração de dados, como tarefas de classificação, agrupamento e associação. É possível observar na figura 2.18 uma sequência de abas que dão acesso à essas tarefas.

Figura 2.18 – Inteface gráfica da função Explorer



Fonte: Elaborado pelo autor.

Para cada tarefa de mineração de dados, existem várias técnicas e algoritmos que podem ser usados na execução dos testes no WEKA. Na classificação, por exemplo, pode-se citar árvore de decisão, *multilayer perceptron* e Naive Bayes. Para a tarefa de agrupamento, o WEKA disponibiliza algoritmos como *Canopy, Cobweb* e EM. Já na tarefa de associação é possível utilizar os algoritmos Apriori, *FilterAssociator* e FP Growth.

2.4 Estudos correlatos

Existem diversos trabalhos relevantes relacionados ao uso de inteligência artificial aplicando técnicas de mineração de dados na predição de ativos da bolsa de valores. Dentre eles, o trabalho de Berenstein (2010) utilizou o processo de KDD para descobrir conhecimento através de indicadores de um conjunto de ativos da bolsa de valores brasileira, utilizando o *software* WEKA na mineração de dados, aplicando

técnicas de mineração de dados para criar regras para indicar um possível comportamento dos ativos selecionados.

Berenstein (2010) utilizou dados de 17 papéis negociados na bolsa de valores brasileira, optando por usar os indicadores de oscilação e negociações realizadas diariamente, sendo que foram incluídos os índices Bovespa e Dow Jones para fim de enriquecimento dos dados. Os dados históricos referentes ao período de janeiro de 2009 até fevereiro de 2010 foram extraídos do site da BM&FBovespa (atualmente conhecida como B3), limpos e transformados de acordo com a estrutura exigida pelo *software* WEKA. O autor aplicou técnicas de árvore de decisão e regras de classificação, sendo que foram usados os algoritmos j48 na árvore de decisão, JRip e PART nas regras de classificação.

Em seu trabalho, Giacomel (2016) propõe modelos *ensembles* baseados em redes neurais para previsão de séries temporais no âmbito da bolsa de valores. São dois *ensembles* diferentes, que são pensados de acordo com o perfil do investidor: o primeiro visa investidores com perfil moderado e o segundo é para quem se propõe a arriscar mais, com o perfil agressivo. O objetivo de ambos é classificar saídas através do uso de redes neurais aplicada em séries temporais, predizendo tendências de alta e queda dos preços dos ativos, e então com esses resultados, auxiliar o investidor a decidir entre comprar ou vender. O autor utiliza para criação das redes neurais o *framework* Encong, que pode ser encontrado para as linguagens de programação Java, C++ e .Net.

Marangoni (2010) tem como objetivo em seu trabalho, a criação de um modelo utilizando redes neurais para predição do preço de fechamento de ações, visando prever o preço da ação da Petrobrás PETR4. Para construção da rede neural o autor utilizou o *software* MATLAB r2009a e um programa escrito pelo próprio autor na linguagem C++, que foi utilizado para construir matrizes a partir de uma planilha fornecida como entrada. Os dados utilizados no trabalho foram obtidos pelo *software* Economática, extraindo os preços diários da ação PETR4 e Índice da Bolsa de Valores de São Paulo (IBOVESPA) do período janeiro de 1999 a maio de 2010. Para criação da tabela, além do uso do fechamento da PETR4 e IBOVESPA e volume de negociações da PETR4, foram usadas entradas calculadas a partir de dados existentes.

3 MATERIAIS E MÉTODOS

Essa seção tem como objetivo descrever os materiais e métodos utilizados na construção desse trabalho, bem como suas etapas e processos na realização das atividades do trabalho.

3.1 Métodos

Segundo Prodanov e Freitas (2013), pesquisa é um processo que busca por respostas para solucionar um problema, por meio de uma sequência de ações, e ao final essas respostas podem ser encontradas ou não. Então, o ato de pesquisar refere-se a busca por um conhecimento que não se tem, mas que, por algum motivo, é importante para melhorar o ambiente de tecnologia, ciência ou bem-estar.

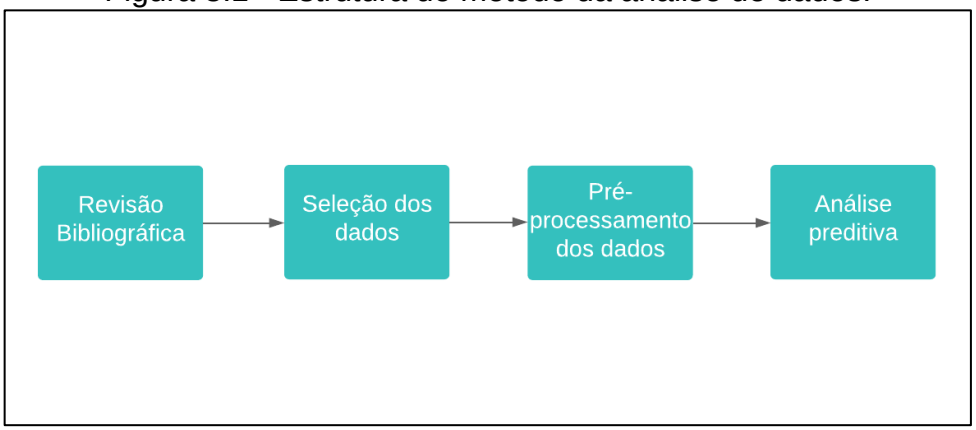
Existem diversas formas para se classificar uma pesquisa cientifica, as principais são por meio de sua finalidade, objetivos, abordagem e pelos procedimentos técnicos. Este trabalho, do ponto de vista da sua finalidade, busca produzir conhecimento a fim de solucionar problemas específicos na prática, caracterizando sua finalidade em uma pesquisa aplicada (PRODANOV; FREITAS, 2013).

Do ponto de vista de seus objetivos, este trabalho pode ser classificado como uma pesquisa exploratória e descritiva, pois em um primeiro momento, foi realizado um levantamento bibliográfico, a fim de proporcionar uma maior familiaridade sobre bolsa de valores e mineração de dados. Em um segundo momento, são coletados dados da bolsa de valores e realizados experimentos com mineração de dados com propósito de estabelecer relações entre as variáveis obtidas, e então são descritos os resultados obtidos. A abordagem desse trabalho é do tipo quantitativa, pois são realizadas análises dos resultados com base em técnicas estatísticas (GIL, 2017).

Esse estudo é classificado como pesquisa bibliográfica e experimental em termos de procedimentos técnicos, sendo que foi elaborado a partir de material bibliográfico já publicado, como livros, teses e monografias, em que tem como objetivo obter conhecimento a partir de material já escrito sobre mineração de dados e bolsa de valores. A manipulação de variáveis para observar os resultados, é uma característica de pesquisas experimentais. Nesse trabalho, são manipuladas variáveis

da bolsa de valores com propósito de analisar os resultados obtidos na mineração de dados (PRODANOV; FREITAS, 2013).

Figura 3.1 - Estrutura do método da análise de dados.



Fonte: Elaborado pelo autor.

A execução desse trabalho foi dividida em quatro etapas, conforme ilustrado na figura 3.1, que se inicia na revisão bibliográfica, passando pela seleção dos dados, pré-processamento dos dados e por fim análise preditiva. Na primeira etapa, na revisão bibliográfica, buscou-se por trabalhos similares ao tema escolhido, livros relacionados ciência de dados, bolsa de valores, DCBD e técnicas de mineração de dados. Foram realizados estudos com objetivo de obter conhecimentos relacionados ao tema, para definir qual a melhor estratégia para resolver o problema proposto.

Na segunda etapa, foram feitas buscas por dados da bolsa de valores, e selecionou-se dados históricos diários da ação da Petrobrás (PETR4). Utilizando a plataforma a *Yahoo Finance*, foram coletados dados históricos da PETR4, a plataforma disponibilizou um arquivo no formato *comma-separated values* (CSV) com os dados do período entre 01/01/2019 e 31/12/2019.

A etapa de pré-processamento consiste na limpeza e transformação dos dados, na terceira etapa foram realizados procedimentos para estruturar, organizar e limpar os dados. E então, na quinta etapa, é realizada a análise preditiva, em que os dados são processados e são aplicadas técnicas de mineração de dados para predição dos preços da ação PETR4, nessa etapa, são aplicadas técnicas de árvore de decisão e redes neurais por meio do *software* WEKA, em cada técnica são utilizados três arquivos com estruturas diferentes e são analisados índices que mostram o nível de acertos obtidos na classificação usando cada técnica.

3.2 Materiais

São utilizados nesse trabalho, o *software* WEKA, dados coletados da plataforma *Yahoo Finance* e um notebook da marca Dell com as seguintes configurações:

- Modelo Dell *Inspiron* 7572;
- Windows 10 *Home*;
- Processador Inter® Core™ i5-8250U CPU 1.60 GHz – 1.80 GHz;
- SSD 240 GB;
- 8 GB RAM.

4 RESULTADOS E DISCUSSÃO

Esta seção tem como objetivo apresentar os resultados obtidos ao longo desse trabalho. Foram realizados testes empregando algoritmos de árvore de decisão e redes neurais e nos testes realizados com árvore de decisão foi utilizado o algoritmo J45, que é uma variação do algoritmo C.45. Já nos testes realizados com redes neurais é utilizado *multilayer perceptron*, apresentado no capítulo 2.

Todos os testes foram feitos de acordo com duas opções de testes fornecidas pelo *software* WEKA. A primeira opção de teste é a *Percentage split,* que consiste em dividir o *dataset,* fornecido previamente, em duas bases de dados, reservando um percentual pré-definido dos dados para o treinamento e o restante para os testes. Na segunda opção de teste, toda a base de dados é utilizada tanto no treinamento, quanto nos testes. Esse tipo de teste é ativado pela opção *Use training set.*

4.1 Seleção e pré-processamento dos dados

Existem diversas fontes para esse tipo de dados de preços de compra e venda de ações e muitas plataformas digitais disponibilizam dados referentes a bolsa de valores. Uma delas é a plataforma *Yahoo Finance,* que disponibiliza dados de todos as ações negociadas na bolsa brasileira, além de indicadores de ações negociadas em bolsas internacionais. A *Yahoo Finance* disponibiliza os seguintes dados históricos de ações: data, abertura, máximo, mínimo, fechamento, fechamento ajustado e volume. É possível escolher o período no qual se deseja visualizar os dados históricos, além da plataforma disponibilizar o *download* de arquivos no formato CSV com os dados.

Para esse trabalho, foram coletados no *Yahoo Finance* dados históricos diários da ação PETR4 da Petrobrás, do período entre 02/01/2019 e 30/12/2019 por meio do *download* de um arquivo CSV. O arquivo baixado tem o formato apresentado na figura 4.1, em que cada coluna é separada por vírgula e respectivamente representa os valores de cada um dos atributos. Um algoritmo escrito pelo autor, que pode ser visto no apêndice A, foi usado para organizar os dados, além de correções manuais usadas na limpeza de dados inconsistentes e transformação dos dados.

Figura 4.1 – Estrutura arquivo com dados da ação PETR4.SA.

```
Date,Open,High,Low,Close,Adj Close,Volume
2019-01-02,22.549999,24.200001,22.280001,24.059999,22.526630,104534800
2019-01-03,23.959999,24.820000,23.799999,24.650000,23.079029,95206400
2019-01-04,24.850000,24.940001,24.469999,24.719999,23.144567,72119800
2019-01-07,24.850000,25.920000,24.700001,25.110001,23.509716,121711900
2019-01-08,25.400000,25.420000,24.770000,24.959999,23.369276,68761800
2019-01-09,25.299999,25.500000,25.150000,25.480000,23.856134,70177600
2019-01-10,25.260000,25.410000,25.070000,25.260000,23.650152,54763700
2019-01-11,25.150000,25.190001,24.840000,24.990000,23.397362,53980700
2019-01-14,24.820000,25.090000,24.660000,24.850000,23.266285,46577300
2019-01-15,24.830000,25.120001,24.700001,24.830000,23.247559,53731500
2019-01-16,24.799999,24.900000,24.660000,24.820000,23.238197,40425000
                                  ...
                                  ...
                                  ...
```

Fonte: Elaborado pelo autor.

Com os dados coletados, decidiu-se trabalhar apenas com os dados do fechamento da ação PETR4. Foram realizados três experimentos, em que os preços do fechamento foram estruturados de maneira que ficassem agrupados em uma linha, blocos sequenciados dos últimos cinco, dez e vinte últimos dias, conforme o experimento realizado. A expressão 4.1 ilustra o formato em que são estruturados os cinco últimos dias. Essa estrutura foi submetida a uma classificação, na qual, foi levado em conta o dia D_k (dia atual) e o dia D_{k-1} (dia anterior).

$$D_{k-4}, D_{k-3}, D_{k-2}, D_{k-1}, D_k \tag{4.1}$$

Então foi criado um atributo a mais, que pode assumir duas situações, a classe *buy* assume que o preço do dia anterior é menor que o dia atual, ou seja, haverá uma valorização no preço de um dia para o outro, então deve-se comprar para se obter lucro, a classe *sell* assume o contrário, pois o dia anterior tem valor maior que o dia atual, então deve-se vender para não ter prejuízo. Por fim, um arquivo no formato ARFF para cada experimento foi gerado a partir do resultado desse pré-processamento. A figura 4.2 ilustra o formato final do arquivo gerado com 5 dias.

Figura 4.2 – Estrutura final do arquivo ARFF com 5 dias de fechamento.

```
@relation PETR4.SA

@attribute Dia1 numeric
@attribute Dia2 numeric
@attribute Dia3 numeric
@attribute Dia4 numeric
@attribute Dia5 numeric
@attribute Action {buy,sell}

@data
24.059999, 24.650000, 24.719999, 25.110001, 24.959999, sell
24.650000, 24.719999, 25.110001, 24.959999, 25.480000, buy
24.719999, 25.110001, 24.959999, 25.480000, 25.260000, sell
25.110001, 24.959999, 25.480000, 25.260000, 24.990000, sell
24.959999, 25.480000, 25.260000, 24.990000, 24.850000, sell
                                ...
                                ...
                                ...
```

Fonte: Elaborado pelo autor.

4.2 Experimentos utilizando árvore de decisão e redes neurais com 5 dias

4.2.1 Árvore de decisão

Os experimentos utilizando árvore de decisão realizados com cinco dias foram feitos a partir de um arquivo ARFF, que possui 243 registros, e desses registros, 124 estão classificados como *buy* e 119 estão classificados como *sell*. No primeiro teste de árvore de decião, foi utilizado o modelo de teste *Percentage split*, que dividiu o *dataset* em 66% para o treinamento e 34% para os testes.

Figura 4.3 – Resultados árvore de decisão com cinco dias – *percentage split.*

```
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          37               44.5783 %
Incorrectly Classified Instances        46               55.4217 %
Kappa statistic                          0
Mean absolute error                      0.5014
Root mean squared error                  0.5015
Relative absolute error                100.0033 %
Root relative squared error            100.0041 %
Total Number of Instances               83

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,000    0,000    ?          0,000    ?          ?        0,500     0,554     buy
                 1,000    1,000    0,446      1,000    0,617      ?        0,500     0,446     sell
Weighted Avg.    0,446    0,446    ?          0,446    ?          ?        0,500     0,506

=== Confusion Matrix ===

  a  b   <-- classified as
  0 46 |  a = buy
  0 37 |  b = sell
```

Fonte: Elaborado pelo autor.

Esse experimento gerou uma árvore com 43 nós, em que 22 são nós folhas. A figura 4.3 mostra os resultados obtidos e na matriz de confusão nota-se que o modelo foi capaz de classificar 100% corretamente quando exposto aos registros de classe *sell*, porém não obteve o mesmo êxito quando exposto aos registros de classe *buy,* em que errou todas as previsões, classificando todos os registros na classe *sell* quando deveria classificar como *buy*. O resultado mostrou-se ruim, pois nesse modelo de testes, se obteve apenas 44,58% de acurácia.

No segundo experimento foi utilizado o modelo de teste em que se usa o *dataset* inteiro tanto para treinamento quanto para os testes. Nesse modelo de testes, o experimento gerou uma árvore com 43 nós, sendo que 22 são nós folhas. O modelo obteve 97,53% de acurácia, acertando 237 classificações das 243 possíveis como ilustrado na figura 4.4. A matriz de confusão mostra que 122 registros são classificados corretamente como *buy*, 115 deles são classificados corretamente como *sell* e outros 6 registros são classificados erroneamente.

Figura 4.4 – Resultados árvore de decisão com cinco dias – treinamento.

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         237               97.5309 %
Incorrectly Classified Instances         6                2.4691 %
Kappa statistic                          0.9506
Mean absolute error                      0.0445
Root mean squared error                  0.1492
Relative absolute error                  8.9087 %
Root relative squared error             29.8475 %
Total Number of Instances              243

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,984    0,034    0,968      0,984    0,976      0,951    0,993     0,991     buy
                 0,966    0,016    0,983      0,966    0,975      0,951    0,993     0,992     sell
Weighted Avg.    0,975    0,025    0,975      0,975    0,975      0,951    0,993     0,991

=== Confusion Matrix ===

   a   b   <-- classified as
 122   2 |   a = buy
   4 115 |   b = sell
```

Fonte: Elaborado pelo autor.

O segundo experimento se mostra mais acurado, embora não se pode dizer que o modelo seja adequado, pois no teste, o modelo é submetido ao mesmo *dataset* em que é treinado, o que pode causar *overfiting,* pois o modelo treinado já conhece os dados de teste. Então um modelo em que não se usa os mesmos dados para treinar e fazer testes, pode ser considerado um modelo com resultados mais aplicáveis ao mundo real.

4.2.2 Redes neurais

Para os experimentos com cinco dias, também foram utilizadas redes neurais. Nesses experimentos, o *dataset* é o mesmo que nos experimentos anteriores. Nesse primeiro experimento, foi realizado um teste dividindo o *dataset* em grupos, sendo que parte dos registros, 66% dos registros, foi destinado para treinamento e o restante separado para testes.

Figura 4.5 – Resultados redes neurais com cinco dias – *percentage split.*

```
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          82               98.7952 %
Incorrectly Classified Instances         1                1.2048 %
Kappa statistic                          0.9756
Mean absolute error                      0.0298
Root mean squared error                  0.1019
Relative absolute error                  5.9402 %
Root relative squared error             20.3247 %
Total Number of Instances               83

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1,000    0,027    0,979      1,000    0,989      0,976    1,000     1,000     buy
                 0,973    0,000    1,000      0,973    0,986      0,976    1,000     1,000     sell
Weighted Avg.    0,988    0,015    0,988      0,988    0,988      0,976    1,000     1,000

=== Confusion Matrix ===

  a  b   <-- classified as
 46  0 |  a = buy
  1 36 |  b = sell
```

Fonte: Elaborado pelo autor.

Analisando a figura 4.5, é possível observar que, o primeiro experimento utilizando redes neurais com cinco dias obtém resultados melhores que nos experimentos com árvore de decisão. O experimento mostrou-se ter 98,80% de acurácia, conseguindo acertar 82 dos 83 registros destinados aos testes. A matriz de confusão confirma o resultado, mostrando que o modelo classificou 46 registros corretamente com *buy* e apenas um dos registros foi classificado equivocadamente na mesma categoria, consequentemente, foram classificados corretamente 36 registros como *sell*, não tendo nenhum erro nas classificações da categoria.

O segundo experimento utilizando redes neurais com cinco dias, usou nos testes o mesmo *dataset* utilizado no treinamento. A figura 4.6 expõe os resultados encontrados nesse teste, no qual, é possível observar que o modelo acertou 97,12% dos registros, errando apenas sete registros do conjunto de dados. A matriz de confusão ilustra bem o resultado, em que 123 registros foram classificados como *buy,* sendo que apenas três classificações estavam na categoria errada, outros 120 foram classificados na categoria *sell*, tendo sucesso em 116 classificações.

Figura 4.6 – Resultados redes neurais com cinco dias - treinamento.

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         236               97.1193 %
Incorrectly Classified Instances         7                2.8807 %
Kappa statistic                          0.9424
Mean absolute error                      0.0389
Root mean squared error                  0.1418
Relative absolute error                  7.7851 %
Root relative squared error             28.3707 %
Total Number of Instances              243

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,968    0,025    0,976      0,968    0,972      0,942    0,998     0,998     buy
                 0,975    0,032    0,967      0,975    0,971      0,942    0,998     0,998     sell
Weighted Avg.    0,971    0,029    0,971      0,971    0,971      0,942    0,998     0,998

=== Confusion Matrix ===

   a   b   <-- classified as
 120   4 |   a = buy
   3 116 |   b = sell
```

Fonte: Elaborado pelo autor.

Os resultados com cinco dias dos experimentos utilizando redes neurais, se mostraram melhores que nos experimentos utilizando árvore de decisão com a mesma quantidade de dias. Tanto o experimento testando o modelo com a opção *Percentage spit*, quanto nos testes utilizando a opção *Use training set*, a diferença entre os resultados de redes neurais e árvore de decisão foram enormes, mostrando que, para o modelo com cinco dias o uso de redes neurais se mostra mais promissor que o uso de árvore de decisão.

4.3 Experimentos utilizando árvore de decisão e redes neurais com 10 dias

4.3.1 Árvore de decisão

A partir da etapa de pré-processamento, para esses experimentos, foi gerado um arquivo ARFF com 10 dias. O arquivo contém 238 registros, em que 123 estão na categoria *buy* e 115 são classificados como *sell.* Nesse experimento foi utilizado árvore de decisão, utilizando a função de teste do WEKA *Percentage split*, em que foi reservado 66% dos registros para o treinamento e 34% para os testes.

Figura 4.7 – Resultados árvore de decisão com dez dias – *percentage split*.

```
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          34               41.9753 %
Incorrectly Classified Instances        47               58.0247 %
Kappa statistic                         -0.1457
Mean absolute error                      0.5335
Root mean squared error                  0.5955
Relative absolute error                106.8183 %
Root relative squared error            119.2151 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,326    0,474    0,438      0,326    0,373      -0,151   0,434     0,527     buy
                 0,526    0,674    0,408      0,526    0,460      -0,151   0,434     0,440     sell
Weighted Avg.    0,420    0,568    0,424      0,420    0,414      -0,151   0,434     0,486

=== Confusion Matrix ===

  a  b   <-- classified as
 14 29 |  a = buy
 18 20 |  b = sell
```

Fonte: Elaborado pelo autor.

O primeiro experimento, obteve uma acurácia de 41,98%, em que 34 registros foram classificados corretamente e 47 registros tiveram classificações incorretas, como apresentado na figura 4.7. Analisando a matriz de confusão, é possível observar como o experimento chegou a esses números, dos 32 registros classificados pelo modelo na categoria *buy,* apenas 14 estavam corretos e nas 49 classificações agrupadas na categoria *sell,* 29 delas estavam incorretas, restando apenas 20 corretas.

No segundo experimento com 10 dias, o resultado não foi muito diferente do experimento anterior. Apesar de ter realizado os testes usando o mesmo *dataset* que foi utilizado no treinamento, o modelo não obteve bons resultados, como pode ser observado na figura 4.8, o modelo obteve 51.68% de acurácia, conseguindo êxito em 123 classificações das 238 possíveis. Porém, ao analisar a matriz de confusão, observou-se que os acertos ocorreram somente na classificação da categoria *buy*, ou seja, não houve registros classificados como *sell.*

Figura 4.8 – Resultados árvore de decisão com dez dias – treinamento.

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         123               51.6807 %
Incorrectly Classified Instances       115               48.3193 %
Kappa statistic                          0
Mean absolute error                      0.4994
Root mean squared error                  0.4997
Relative absolute error                 99.9991 %
Root relative squared error            100      %
Total Number of Instances              238

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1,000    1,000    0,517      1,000    0,681      ?        0,500     0,517     buy
                 0,000    0,000    ?          0,000    ?          ?        0,500     0,483     sell
Weighted Avg.    0,517    0,517    ?          0,517    ?          ?        0,500     0,501

=== Confusion Matrix ===

   a   b   <-- classified as
 123   0 |   a = buy
 115   0 |   b = sell
```

Fonte: Elaborado pelo autor.

Analisando os experimentos com 10 dias utilizando árvore de decisão, percebe-se que, ambos tiveram desempenho ruim em relação a sua capacidade de classificar corretamente os registros. Os resultados dos experimentos com 5 dias utilizando árvore de decisão, reforçam a hipótese de que o uso de árvore de decisão não cabe nesse modelo de previsão, pois o experimento utilizando a função de testes *Percentage split* também obteve resultados ruins.

4.3.2 Redes neurais

O mesmo *dataset* dos experimentos anteriores com 10 dias foi usado nos experimentos utilizando redes neurais. O primeiro desses experimentos, foi testado no modelo *Percentage split*, no qual foi destinado 66% do *dataset* para treinamento e o restante para testes.

Figura 4.9 – Resultados redes neurais com dez dias – *percentage split*.

```
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          77               95.0617 %
Incorrectly Classified Instances         4                4.9383 %
Kappa statistic                          0.9009
Mean absolute error                      0.0573
Root mean squared error                  0.1892
Relative absolute error                 11.4794 %
Root relative squared error             37.8766 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,953    0,053    0,953      0,953    0,953      0,901    0,993     0,994     buy
                 0,947    0,047    0,947      0,947    0,947      0,901    0,993     0,992     sell
Weighted Avg.    0,951    0,050    0,951      0,951    0,951      0,901    0,993     0,993

=== Confusion Matrix ===

  a  b   <-- classified as
 41  2 |  a = buy
  2 36 |  b = sell
```

Fonte: Elaborado pelo autor.

O experimento obteve 95,06% de acurácia, em que conseguiu acertar 77 dos 81 registros de teste. Analisando a figura 4.9 observa-se que, a matriz de confusão expressa bem o resultado, em que dos 43 registros classificados como *buy*, 41 estavam corretos e conseguindo êxito na classificação de 36 dos 38 classificados como *sell.*

No segundo experimento com 10 dias utilizando redes neurais, foi utilizado para teste a mesma base de dados de treinamento. Nos resultados exibidos na figura 4.10 é possível observar o alto nível de acerto do modelo, que obteve acurácia de 98,74%, tendo êxito em 235 classificações. Ao analisar a matriz de confusão, nota-se que 126 registros foram classificados como *buy*, sendo que, 123 deles estavam corretos e observando a categoria *sell,* foi possível verificar que 100% das classificações feitas pelo modelo mostraram-se corretas.

Figura 4.10 – Resultados redes neurais com dez dias II

Figura 4.10 – Resultados redes neurais com dez dias – treinamento.

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         235               98.7395 %
Incorrectly Classified Instances         3                1.2605 %
Kappa statistic                          0.9747
Mean absolute error                      0.0323
Root mean squared error                  0.1056
Relative absolute error                  6.4601 %
Root relative squared error             21.1389 %
Total Number of Instances              238

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1,000    0,026    0,976      1,000    0,988      0,975    1,000     1,000     buy
                 0,974    0,000    1,000      0,974    0,987      0,975    1,000     1,000     sell
Weighted Avg.    0,987    0,013    0,988      0,987    0,987      0,975    1,000     1,000

=== Confusion Matrix ===

   a   b   <-- classified as
 123   0 |   a = buy
   3 112 |   b = sell
```

Fonte: Elaborado pelo autor.

Os resultados com 10 dias utilizando redes neurais apresentaram níveis de acurácia que podem ser considerados ótimos. Comparando com os resultados com 5 dias utilizando redes neurais, é possível observar que uso de redes neurais para esse modelo de dados se mostra promissor.

4.4 Experimentos utilizando árvore de decisão e redes neurais com 20 dias

4.4.1 Árvore de decisão

Para os experimentos com 20 dias, foi gerado um arquivo ARFF com 228 registros, em que 117 deles são classificados como *buy* e 111 estão na categoria *sell.* O arquivo foi utilizado no primeiro experimento, no qual foi divido em dois grupos, utilizando a função *Percentage split* do WEKA. O primeiro grupo contou com 66% dos registros, sendo destinado para o treinamento do modelo, e o segundo grupo, com 34% foi reservado para validação.

Figura 4.11 – Resultados árvore de decisão com vinte dias - *percentage split.*

```
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          35               44.8718 %
Incorrectly Classified Instances        43               55.1282 %
Kappa statistic                          0
Mean absolute error                      0.5007
Root mean squared error                  0.5007
Relative absolute error                100.0018 %
Root relative squared error            100.002  %
Total Number of Instances               78

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,000    0,000    ?          0,000    ?          ?        0,500     0,551     buy
                 1,000    1,000    0,449      1,000    0,619      ?        0,500     0,449     sell
Weighted Avg.    0,449    0,449    ?          0,449    ?          ?        0,500     0,505

=== Confusion Matrix ===

  a  b   <-- classified as
  0 43 |  a = buy
  0 35 |  b = sell
```

Fonte: Elaborado pelo autor.

Observando os resultados do primeiro experimento com 20 dias utilizando árvore de decisão mostrados na figura 4.11, observa-se que a acurácia do experimento é de 44,87%, o que se mostra baixo para os padrões aceitáveis. Analisando a matriz de confusão, nota-se que os registros foram classificados apenas na categoria *sell,* no qual 35 dos 78 registros separados para os testes foram classificados corretamente e 43 registros tiveram classificações erradas.

Para o segundo experimento com 20 dias utilizando árvore de decisão, foi utilizado o modelo de testes *Use training set,* utilizando todo o *dataset* para o treinamento e testes. Nos resultados apresentados na figura 4.12, observa-se que a acurácia do modelo é de 52,32%, na qual o modelo consegue acertar 117 dos 228 registros *dataset.* Porém, a matriz de confusão apresentada na figura 4.12, mostra todos os registros foram classificados como *buy,* no qual 111 classificações estavam incorretas.

Figura 4.12 – Resultados árvore de decisão com vinte dias – treinamento.

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         117               51.3158 %
Incorrectly Classified Instances       111               48.6842 %
Kappa statistic                          0
Mean absolute error                      0.4997
Root mean squared error                  0.4998
Relative absolute error                 99.9994 %
Root relative squared error            100      %
Total Number of Instances              228

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1,000    1,000    0,513      1,000    0,678      ?        0,500     0,513     buy
                 0,000    0,000    ?          0,000    ?          ?        0,500     0,487     sell
Weighted Avg.    0,513    0,513    ?          0,513    ?          ?        0,500     0,500

=== Confusion Matrix ===

   a   b   <-- classified as
 117   0 |   a = buy
 111   0 |   b = sell
```

Fonte: Elaborado pelo autor.

Os resultados dos experimentos com 20 dias utilizando árvore de decisão, se mantiveram parecidos com os experimentos anteriores envolvendo árvore de decisão, fazendo com que a hipótese levantada nos experimentos com 10 dias utilizando árvore de decisão ganhe mais força.

4.4.2 Redes neurais

O arquivo ARFF utilizado nos experimentos utilizando redes neurais com 20 dias, foi o mesmo utilizando nos experimentos com árvore de decisão. O primeiro experimento foi realizado utilizando a opção de teste do *software* WEKA *Percentage split,* em que foi destinado para treinamento do modelo 66% do *dataset* e utilizado nos testes 34%.

Figura 4.13 – Resultados redes neurais com vinte dias – *percentage split*

```
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          75               96.1538 %
Incorrectly Classified Instances         3                3.8462 %
Kappa statistic                          0.9225
Mean absolute error                      0.076
Root mean squared error                  0.1872
Relative absolute error                 15.1791 %
Root relative squared error             37.3785 %
Total Number of Instances               78

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,953    0,029    0,976      0,953    0,965      0,923    0,995     0,996     buy
                 0,971    0,047    0,944      0,971    0,958      0,923    0,995     0,994     sell
Weighted Avg.    0,962    0,037    0,962      0,962    0,962      0,923    0,995     0,995

=== Confusion Matrix ===

  a  b   <-- classified as
 41  2 |  a = buy
  1 34 |  b = sell
```

Fonte: Elaborado pelo autor.

O primeiro experimento teve como resultado acurácia de 96,15%, em que o modelo obteve êxito em 75 classificações, em que o número total de classificações possíveis é de 78, como pode ser observado na figura 4.13. Pode-se observar que, a matriz de confusão representa esse resultado, no qual, 42 registros foram classificados como *buy*, sendo que apenas uma dessas classificações estavam incorretas, os registros classificados como *sell* somam 36, em que 34 classificações estavam corretas.

Finalmente, o último experimento em que se usou redes neurais com 20 dias, utilizou a função de testes do WEKA *Use training set*, em que foi utilizado todo *dataset* para treinamento e teste. O experimento mostrou-se impecável, no qual classificou 100% dos registros corretamente, como pode ser observado na figura 4.14. Analisando a matriz de confusão, percebe-se que, 117 registros são classificados como *buy* e 111 deles estão na categoria *sell,* o que corresponde aos 100% de acerto citado anteriormente.

Figura 4.14 – Resultados redes neurais com vinte dias – treinamento.

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         228               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0.0128
Root mean squared error                  0.0283
Relative absolute error                  2.561  %
Root relative squared error              5.6548 %
Total Number of Instances              228

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1,000    0,000    1,000      1,000    1,000      1,000    1,000     1,000     buy
                 1,000    0,000    1,000      1,000    1,000      1,000    1,000     1,000     sell
Weighted Avg.    1,000    0,000    1,000      1,000    1,000      1,000    1,000     1,000

=== Confusion Matrix ===

   a   b   <-- classified as
 117   0 |   a = buy
   0 111 |   b = sell
```

Fonte: Elaborado pelo autor.

Os resultados com 20 dias utilizando redes neurais se mostraram ótimos, conseguindo acurácias acima de 96%, corroborando com os resultados anteriores em que se usa redes neurais, mostrando que são promissores estudos envolvendo redes neurais aplicados na previsão de ativos da bolsa de valores.

4.5 DISCUSSÕES

Os resultados apresentados na seção 4.4 se mostraram melhores quando realizados com redes neurais, como pode ser observado na tabela 4.1, seus níveis de acurácia se mantiveram acima de 95%. Na contramão dos resultados obtidos com redes neurais, os níveis de acurácia obtidos nos experimentos com árvore de decisão se mantiveram abaixo de 52% na maioria dos experimentos.

Os resultados utilizando o mesmo *dataset* para treinamento e para realização dos testes, em sua maioria se obteve bons números, apesar de que, nesse tipo de teste pode ocorrer *overfitting,* esses números podem ser importantes na escolha dos algoritmos a serem usados. O maior índice desse tipo de treinamento foi encontrado no experimento com vinte dias utilizando redes neurais, onde se obteve 100% de

acerto e o menor nível aconteceu no experimento com vinte dias utilizando árvore de decisão, em que se atingiu a acurácia de 51,32% de acerto.

Tabela 4.1 - Resultados dos experimentos realizados.

Experimentos	Árvore de decisão (%)	Redes Neurais (%)
Cinco dias com teste split 66%	44,58	98,80
Cinco dias com teste de treinamento	97,53	97,12
Dez dias com teste split 66%	41,98	95,06
Dez dias com teste de treinamento	51,68	98,74
Vinte dias com teste split 66%	44,87	96,15
Vinte dias com teste de treinamento	51,32	100,00

Fonte: Elaborado pelo autor.

Analisando os experimentos em que foi utilizado a opção de testes *Percentage split,* o melhor resultado apareceu no experimento com cinco dias utilizando redes neurais, em a acurácia chegou a 98,80% de acertos. E o pior resultado nesse modelo de testes, ocorreu no teste de no experimento em que se utilizou árvore de decisão com dez dias, atingindo o nível de 41,98% de acertos dos registros.

Pode-se dizer que o experimento que se mostrou mais eficiente, foi o experimento com vinte dias utilizando redes neurais, pois nos testes utilizando parte dos registros como teste e outra parte para treinamento, se obteve o resultado de 96,15%, sendo confirmado o bom resultado nos testes usando o set de treinamento, em que se obteve 100% de acerto, o que confirma sua eficiência.

5 CONSIDERAÇÕES FINAIS

A tarefa de prever preços de ações é enormemente complicada, pois existem muitas variáveis que influenciam o mercado. Um desastre natural pode empurrar os preços de algumas ações para baixo, assim como um aceno do governo de um projeto para baixar taxas e juros e movimentar o mercado. Prever os preços com alguns desses acontecimentos é um desafio enorme, as vezes impossível. No entanto, já existem várias ferramentas que podem cumprir essa tarefa com um nível de erro baixo.

A aplicação da mineração de dados dentro do processo de descoberta de conhecimento em base de dados, pode ser uma grande aliada de investidores na sua tomada de decisão. A aplicação das técnicas de mineração de dados nesse trabalho mostra que, é possível prever com certo grau de acurácia se o preço futuro de uma ação vai subir ou descer, pelo menos.

Para o cumprimento do objetivo geral, procurou-se satisfazer os itens dos objetivos específicos, os quais foram definidos no capítulo 1 (um) desse trabalho. A conceituação dos temas mercado de ações e mineração de dados, decorreram de uma pesquisa exploratória, em que foram buscados autores conhecidos e trabalhos relevantes relacionados aos temas.

Quanto ao item que trata da aplicação do processo de descoberta de conhecimento, foi realizada uma busca por dados históricos, e foi decidido que os dados fossem baixados do site Yahoo Finance, que disponibiliza diversos tipos de dados do mercado financeiro. A aplicação das etapas foi realizada conforme apresentado no capítulo 3 (três), em que na etapa de mineração de dados, foram utilizadas técnicas de redes neurais e árvore de decisão.

A análise dos resultados obtidos nos experimentos com WEKA, foi feita por meio de uma análise descritiva dos dados, em que são apresentados. Então, foi realizada uma análise da capacidade de predição proveniente dos seus resultados de maneira comparativa.

Com base em toda pesquisa desse trabalho, chegou-se à conclusão de que, a aplicabilidade da mineração de dados no mundo dos investimentos é possível, se a tarefa utilizada for escolhida de maneira correta. No caso desse trabalho, o uso de árvore de decisão não se mostrou eficiente como de fato ocorreu em outros trabalhos

correlatos. Por outro lado, analisando os modelos que utilizaram redes neurais, todos os experimentos obtiveram resultados acima adequado.

5.1 Trabalhos futuros

Para trabalhos futuros relacionados a essa pesquisa, recomenda-se que:

- Incluir indicadores técnicos;
- Utilizar tendências;
- Realizar testes com *dataset* maiores;
- Utilizar programação, com objetivo de ter mais controle sobre os dados.

REFERÊNCIAS

AGOSTINI, Michel**. Estudo comparativo entre as ferramentas weka e sas no processo de descoberta de informações.** 2017. 55 f. Monografia (Especialização) - Curso de Especialização em Banco de Dados, Universidade Federal de Mato Grosso, Cuiabá, 2017.

ASSAF NETO, Alexandre. **Mercado financeiro**. 14. ed. São Paulo: Atlas, 2018.

BERENSTEIN, Marcelo. **Uso de mineração de dados na bolsa de valores.** 2010. 95 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade do Vale do Itajaí, Itajaí, 2010.

BRATTI, Diogo. **Sistema informatizado para auxiliar o pequeno investidor na bolsa de valores**. 2009. 398 f. TCC (Graduação) - Curso de Ciências da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2009.

CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. **Introdução à mineração de dados:** conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.

COMISSÃO DE VALORES MOBILIÁRIOS (org.). **Mercado de valores mobiliários brasileiro.** 4. ed. Rio de Janeiro: Comissão de Valores Mobiliários, 2019.

D'ÁVILA, Mariana Zonta. **Bolsa conquista 1,5 milhão de novos investidores em 2020, um aumento de 92% no ano. 2021.** Disponível em: https://www.infomoney.com.br/onde-investir/bolsa-conquista-15-milhao-de-novos-investidores-em-2020-um-aumento-de-92-no-ano/. Acesso em: 29 maio 2021.

DEBASTIANI, Carlos Alberto. **Candlestick**: Um método para ampliar lucros na bolsa de valores. São Paulo: Novatec, 2007.

ELIAS, Juliana. **De 354 ações da B3, só 21 estão no azul – e é mais sorte que oportunidade**. 2020. Disponível em:

https://www.cnnbrasil.com.br/business/2020/03/31/de-354-acoes-da-b3-so-21-estao-no-azul-e-e-mais-sorte-que-oportunidade. Acesso em: 09 abr. 2021

FAYYAD, U.; PIATESKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. **Advances in knowledge discovery and data mining**. 1996. Cambridge: MIT Press, 1996. 560p.

FOGAÇA, André. **Bolsa de valores para leigos.** São Paulo: Guiainvest, 2015.

GIACOMEL, Felipe dos Santos. **Um Método Algorítmico para Operações na Bolsa de Valores Baseado em Ensembles de Redes Neurais para Modelar e Prever os Movimentos dos Mercados de Ações**. 2016. 92 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.

HAYKIN, Simon. **Redes Neurais:** princípios e prática. 2. ed. Porto Alegre: Bookman, 2001.

LEMOS, Flávio. **Análise técnica dos mercados financeiros:** um guia completo e definitivo dos métodos de negociação de ativos. São Paulo: Saraiva, 2016.

MARANGONI, Pedro Henrique. **Redes Neurais Artificiais para Previsão de Séries Temporais no Mercado Acionário.** 2010. 80 f. TCC (Graduação) - Curso de Ciências Econômicas, Universidade Federal de Santa Catarina, Florianópolis, 2010.

MOURA, Karina. **Ciclo de vida dos dados**: kdd process. 2019. Disponível em: https://medium.com/@kvmoura/kdd-process-9b8e3062142. Acesso em: 18 abr. 2021.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico:** métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo

Amburgo: Universidade Feevale, 2013. Disponível em: https://www.feevale.br/institucional/editora-feevale/metodologia-do-trabalho-cientifico---2-edicao. Acesso em: 11 abr. 2021.

ROQUE, Reginaldo do Carmo. **Estudo sobre a empregabilidade da previsão do índice BOVESPA usando redes neurais artificiais**. 2009. 102 f. TCC (Graduação) - Curso de Engenharia Eletrônica e de Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009. Disponível em: https://pantheon.ufrj.br/handle/11422/7567. Acesso em: 02 mar. 2021.

RUSSELL, Stuart; NORVIG, Peter**. Inteligência artificial.** 3. ed. Rio de Janeiro: Elsevier, 2013.

SCHMELZER, Ron. **Unleashing the real power of data**. 2020. Disponível em: https://www.forbes.com/sites/cognitiveworld/2020/02/06/unleashing-the-real-power-of-data/?sh=4ab84377389d. Acesso em: 17 abr. 2021.

SILVA, Fabrício Machado da et al. **Inteligência Artificial**. Porto Alegre: Sagah, 2019.

SILVA, Leandro Augusto da; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução à mineração de dados**: com aplicações em r. Rio de Janeiro: Elsevier, 2016.

SILVA, Marcelino Pereira dos Santos; **Mineração de dados:** conceitos, aplicações e experimentos com weka. In: Simpósio de informática do CEFET-PI, 6, 2008, Teresina. **Minicurso,** Teresina: CEFET-PI. P. 1-20.

APÊNDICE A – Código Java para estruturar dados da ação petr4.

```
package petr4;

import java.io.BufferedReader;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.PrintWriter;
import java.util.ArrayList;

/**
 *
 * @author Wanderson Bleiner
 */
public class petr4 {



    public static void main(String[] args) {
        // TODO code application logic here

        String                      arqEntrada                      =
"C:\\Users\\blein\\OneDrive\\Documentos\\NetBeansProjects\\novopetr4
\\src\\novopetr4\\PETR4.SA_entrada.csv";
        String                       arqSaida                       =
"C:\\Users\\blein\\OneDrive\\Documentos\\NetBeansProjects\\novopetr4
\\src\\novopetr4\\PETR4_saida.arff";
        String array[] = new String[7];
        ArrayList<String> aux = new ArrayList();;
        double n1, n2;
        String linha;

        try {
```

```
        FileReader arq = new FileReader(arqEntrada);
        BufferedReader lerArq = new BufferedReader(arq);
        FileWriter arq1 = new FileWriter(arqSaida);
        PrintWriter gravarArq = new PrintWriter(arq1);

        linha = lerArq.readLine();

        while (linha != null) {
            array = linha.split(",");
            aux.add(array[4]);
            linha = lerArq.readLine();
        }
        int j;
        for (int i = 0; i < aux.size() - 10; i++) {
            j = i;
            for (; j < i + 10; j++) {
                gravarArq.printf("%s, ", aux.get(j));

            }

            n1 = Double.parseDouble(aux.get(j - 1));
            n2 = Double.parseDouble(aux.get(j - 2));

            if (n2 < n1) {
                gravarArq.println("buy");
            } else {
                gravarArq.println("sell");
            }
            gravarArq.flush();
        }
        gravarArq.close();

    } catch (Exception e) {
```

```
            System.out.println(e);
        }
    }
}
```

Fonte: Elaborado pelo autor.

ANEXO A – Termo de publicação de produção acadêmica.

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
PRÓ-REITORIA DE GRADUAÇÃO

Av. Universitária, 1069 ● Setor Universitário
Caixa Postal 86 ● CEP 74605-010
Goiânia ● Goiás ● Brasil
Fone: (62) 3946.1021 I Fax: (62) 3946.1397
www.pucgoias.edu.br I prograd@pucgoias.edu.br

ANEXO I
APÊNDICE ao TCC

Termo de autorização de publicação de produção acadêmica

O estudante Wanderson Bleiner Coelho de Souza do Curso de Ciência da Computação, matrícula 2017.2.0028.0054-6, telefone: (62) 98539-1251, e-mail wanderson.bleiner@gmail.com, na qualidade de titular dos direitos autorais, em consonância com a Lei nº 9.610/98 (Lei dos Direitos do Autor), autoriza a Pontifícia Universidade Católica de Goiás (PUC Goiás) a disponbilizar o Trabalho de Conclusão de Curso intitulado Mineração de dados aplicada a previsão de preços de ações utilizando o Weka gratuitamente, sem ressarcimento dos direitos autorais, por 5 (cinco) anos, conforme permissões do codumento, em meio eletrônico, na rede mundia de computadores, no formato especificado (Texto(PDF); Imagem (GIF ou JPEG); Som (WAVE, MPEG, AIFF, SND); Vídeo (MPEG, MWV, AVI, QT); outros), específicos da área para fins de leitura e/ou impressao pela internet, a título de divulgação da produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 15 de junho de 2021

Assinatura do autor: Wanderson Bleiner Coelho de Souza

Nome completo do autor: Wanderson Bleiner Coelho de Souza

Assinatura do professor – orientador: Sibelius Lellis Vieira

Nome completo do professor – orientador: Sibelius Lellis Vieira