



**UNIVERSIDADE DO ESTADO DA BAHIA**  
**DEPARTAMENTO DE CIÊNCIAS EXATAS E DA TERRA**  
**CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO**

**LUCAS FREITAS SILVA**

**PREVISÃO DE SÉRIES TEMPORAIS NO MERCADO FOREX UTILIZANDO  
MÉTODOS DE MINERAÇÃO DE DADOS BASEADO EM ESTADOS  
TERMODINÂMICOS: UMA ABORDAGEM EMPÍRICO-COMPARATIVA.**

**SALVADOR**

**2018**

LUCAS FREITAS SILVA

PREVISÃO DE SÉRIES TEMPORAIS NO MERCADO FOREX UTILIZANDO MÉTODOS  
DE MINERAÇÃO DE DADOS BASEADO EM ESTADOS TERMODINÂMICOS: UMA  
ABORDAGEM EMPÍRICO-COMPARATIVA.

Monografia apresentada ao curso de Sistemas de Informação do Departamento de Ciências Exatas e da Terra da Universidade do Estado da Bahia - UNEB, como requisito parcial à obtenção do grau de bacharel em Sistemas de Informação.

Área de Concentração: Sistemas de Informação

Linha de Pesquisa: Inteligência Artificial

Orientador: PhD. Diego Gervasio Frías Suárez

SALVADOR

2018

FICHA CATALOGRÁFICA  
Sistema de Bibliotecas da UNEB  
Dados fornecidos pelo autor

S586p

Silva, Lucas Freitas

Previsão de séries Temporais no mercado Forex utilizando métodos de mineração de dados baseado em estados termodinâmicos: Uma abordagem empírico-comparativa. / Lucas Freitas Silva.-- Salvador, 2018.

80 fls.

Orientador(a): Diego Gervasio Frías Suárez.

Inclui Referências

TCC (Graduação - Sistemas de Informação) - Universidade do Estado da Bahia. Departamento de Ciências Exatas e da Terra. Câmpus I. 2018.

1.Mineração de Dados. 2.Séries Temporais. 3.Padrões de Análise Técnica.. 4.Inteligência Artificial..

CDD: 004

LUCAS FREITAS SILVA

PREVISÃO DE SÉRIES TEMPORAIS NO MERCADO FOREX UTILIZANDO MÉTODOS  
DE MINERAÇÃO DE DADOS BASEADO EM ESTADOS TERMODINÂMICOS: UMA  
ABORDAGEM EMPÍRICO-COMPARATIVA.

Monografia apresentada ao curso de Sistemas de Informação do Departamento de Ciências  
Exatas e da Terra da Universidade do Estado da Bahia - UNEB, como requisito parcial à  
obtenção do grau de bacharel em Sistemas de Informação.

Área de Concentração: Sistemas de Informação

Linha de Pesquisa: Inteligência Artificial

Aprovada em: 29 de novembro de 2018

BANCA EXAMINADORA

---

PhD. Diego Gervasio Frías Suárez (Orientador)  
Universidade do Estado da Bahia - UNEB

---

Dr. Leandro Santos Coelho de Souza  
Universidade do Estado da Bahia - UNEB

---

Dr. Hugo Saba Pereira Cardoso  
Universidade do Estado da Bahia - UNEB

*Dedico este trabalho à minha mãe, Maria  
Marta. E a minha vó Maria Amelia.*

## AGRADECIMENTOS

Agradeço à minha mãe, por todos os momentos dedicados a mim e todo trabalho árduo, não deixando faltar nada e sempre me guiando e educando com muito amor e carinho para eu alcançar o melhor caminho.

À minha vó, por todas as orações e carinho incondicional.

Agradeço a Deus por me dar força saúde e sabedoria para superar percalços enfrentados.

A todos os meus professores que contribuíram para a minha formação. Em especial ao professor Leandro Coelho que acreditou no meu potencial, mostrando o caminho da ciência através do seu grupo de pesquisa PIMAT. Também ao meu orientador Diego Frias, que me guiou durante todo processo do meu trabalho de conclusão de curso, dando todo suporte necessário, realizando correções e dando broncas quando preciso, mostrando-se presente em todos os momentos.

Aos amigos que fiz ao longo dessa jornada acadêmica, em especial Ana Cecília, Augusto Lima, Stephanie Lima, Claudia Elizabete, Thamires Asenate, Camila Sa e Felipe Andrade, por todos os momentos que passamos juntos, no amor e na dor.

Aos meus amigos que me acompanham desde o ensino médio, Anderson Serrado, Gisele Araújo e Valdete Ribeiro, que me aturam até então.

E agradeço a todos que direta ou indiretamente colaboraram para minha formação

*"Em um estado sombrio nós nos encontramos...  
um pouco mais de conhecimento iluminar nosso  
caminho pode."*

*(Yoda.)*

## RESUMO

Séries Temporais apresentam-se como um grande desafio para a ciência. Um ambiente que fornece séries de alta complexidade e que apresenta importância econômica, histórica e política é o mercado Forex. O mercado Forex é um mercado de câmbio de divisas em âmbito internacional. Diversos *players* importantes, tais como governo e bancos operam neste mercado, atualmente há a possibilidade de usuários comuns operarem neste mercado através de corretoras. Entretanto para operar neste mercado é necessário tomar algumas precauções e ter em mente que isso não é uma tarefa simples, pois o mercado já está munido de estratégias existentes e amplamente divulgadas na internet, assim cada vez mais se faz necessário a busca de novas soluções/estratégias que nos tragam confiança e estabilidade para operar neste mercado, buscando taxas de sucesso cada vez maiores. Sendo assim, o presente trabalho traz uma nova estratégia de operação desde mercado e tem como objetivo modelar com base na análise markoviana os estados termodinâmicos, em forma de markov de segunda ordem. Posteriormente realizar a predição das séries temporais oriundas do Forex com base nas técnicas de mineração de dados buscando viabilizar e transformar este modelo em uma aplicação capaz de prever corretamente os estados e operar automaticamente neste mercado.

**Palavras-chave:** Mineração de dados. Séries Temporais. Padrões de Análise Técnica.

## **ABSTRACT**

Time series present themselves as a major challenge to science. An environment that provides series of high complexity and that presents economic, historical and political importance is the Forex market. The Forex market is an international currency exchange market. Several important players, such as government and banks operate in this market, currently there is the possibility of ordinary users operating in this market through brokerage firms. However, to operate in this market, it is necessary to take some precautions and keep in mind that this is not a simple task, since the market is already equipped with existing strategies and widely disseminated on the Internet, so it is increasingly necessary to search for new solutions/strategies that bring us the confidence and stability to operate in this market, seeking ever greater success rates. Thus, the present work brings a new strategy of operation from the market and aims to model based on the Markovian analysis the thermodynamic states of the Forex market. Subsequently perform the prediction of the time series derived from Forex based on the techniques of data mining seeking to make feasible and transform this model into an application capable of correctly predicting the states and automatically operate in this market.

**Keywords:** FOREX. Data Mining. Time series. Technical Analysis.

LUCAS FREITAS SILVA

**Previsão de séries no mercado Forex utilizando  
métodos de mineração de dados baseado em  
estados termodinâmicos.**

Salvador

2018

# Lista de ilustrações

Figura 1 – Exemplos de moedas negociadas no Forex . . . . .	18
Figura 2 – Representação dos candlesticks do Forex . . . . .	20
Figura 3 – Exemplo de uma série temporal determinística genérica. . . . .	23
Figura 4 – Exemplo série temporal estocástica do Forex. . . . .	24
Figura 5 – Exemplo dos 13 estados e suas relações. . . . .	29
Figura 6 – Ilustração do estado acessível através de um exemplo de jogo cara ou coroa . . . . .	30
Figura 7 – Estados transientes e recorrentes. . . . .	31
Figura 8 – Estados transientes e recorrentes. . . . .	33
Figura 9 – Exemplo do processo KDD. . . . .	35
Figura 10 – Ilustração do modelo. . . . .	40
Figura 11 – Ilustração do roteiro metodológico. . . . .	42
Figura 12 – Ilustração da estratégia de validação. . . . .	43
Figura 13 – Ilustração da etapa do MetaTrader. . . . .	44
Figura 14 – Ilustração da etapa base de dados - Excel. . . . .	45
Figura 15 – Ilustração dos estados e transições. . . . .	47
Figura 16 – Ilustração da grafo orientado. . . . .	47
Figura 17 – Markov primeira Ordem. . . . .	48
Figura 18 – Markov segunda ordem. . . . .	49
Figura 19 – Ilustração de um recorte dos dados. . . . .	49
Figura 20 – Ilustração da etapa do Weka. . . . .	50
Figura 21 – Exemplo de um arquivo Arff genérico. . . . .	52
Figura 22 – Exemplo de uma árvore plotada com o algoritmo j48. . . . .	53
Figura 23 – Representação da Abordagem Formal. . . . .	56
Figura 24 – Gráfico de tendência Downtrend Concavo . . . . .	57
Figura 25 – Gráfico de tendência Downtrend Convexo. . . . .	58
Figura 26 – Gráfico de tendência Uptrend Convexo. . . . .	58
Figura 27 – Gráfico de tendência Uptrend Concavo. . . . .	58
Figura 28 – Conceito da abordagem hibrida. . . . .	59
Figura 29 – Porcentagem da quantidade de instâncias da série temporal - Junho 2018. . . . .	61
Figura 30 – Porcentagem de classes de acordo com os limiares pré definidos - Junho 2018. . . . .	61

Figura 31 – Porcentagem da quantidade de instâncias da série temporal - Outubro 2018. . . . .	62
Figura 32 – Porcentagem de classes de acordo com os limiares pré definidos - Outubro 2018. . . . .	62
Figura 33 – Ilustração do funcionamento do classificador. . . . .	63
Figura 34 – Porcentagens do valores preditos utilizando os <i>datasets</i> de teste. . . . .	67
Figura 35 – Evolução da acurácia da abordagem pragmática. . . . .	68
Figura 36 – Porcentagens do valores preditos utilizando os <i>datasets</i> de teste. . . . .	71
Figura 37 – Resultado percentual da análise estatística da abordagem híbrida. . . . .	73

# Lista de tabelas

Tabela 1 – Estados identificados nas séries temporais do Forex . . . . .	28
Tabela 2 – Técnica de classificação e algoritmos de aprendizado de máquina utilizados. . . . .	55
Tabela 3 – Estados-Transições validos para a realização de operações. . . . .	56
Tabela 4 – Resultado experimento formal . . . . .	64
Tabela 5 – Matriz de confusão . . . . .	65
Tabela 6 – Resultado dos coeficientes Kappa e sua interpretação. . . . .	66
Tabela 7 – Resultado das acurácias do experimento pragmático. . . . .	68
Tabela 8 – Resultado das acurácias do experimento pragmático. . . . .	69
Tabela 9 – Matriz de confusão do algoritmo ZeroR . . . . .	69
Tabela 10 – Resultado dos coeficientes Kappa e sua interpretação. . . . .	70
Tabela 11 – Contabilidade dos Atrasos - Abordagem Formal . . . . .	72
Tabela 12 – Contabilidade dos Atrasos - Abordagem Pragmática . . . . .	72

# Lista de abreviaturas e siglas

EUR	Euro
FOREX	Foreign Exchange
MQL4	<i>MetaQuotes4</i>
MT4	<i>MetaTrader4</i>
PIP	<i>Percentage In Point</i>
USD	Dólar
KDD	Knowledge Discovery in Databases.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>14</b>
<b>2</b>	<b>Mercado Forex</b>	<b>18</b>
2.1	Características	19
2.2	Séries temporais do Forex	20
<b>3</b>	<b>Modelagem de Séries Temporais</b>	<b>21</b>
3.1	Séries Temporais	21
3.1.1	Características e formação.	22
3.1.2	Processo estocástico e estacionariedade	23
3.1.3	Aplicações das séries temporais e importância	24
3.2	ARIMA	25
3.3	Redes Neurais	26
3.4	Outros	26
3.5	Abordagem proposta no trabalho	27
3.5.1	Modelo Termodinâmico dos Estados	27
3.5.2	Cadeias de Markov	28
<b>4</b>	<b>Mineração de Dados.</b>	<b>34</b>
4.1	Descoberta de Conhecimento em Banco de Dados	35
4.2	Mineração de dados aplicadas as séries temporais	36
4.3	Técnicas de Mineração de dados	36
4.3.1	Classificação.	37
4.3.2	Associação	37
4.3.3	Agrupamento	38
4.4	Técnica proposta no trabalho	38
<b>5</b>	<b>Realização da previsão</b>	<b>40</b>
5.1	Metodologia	41
5.2	Modelo	41
5.3	Revisão Sistemática e Literatura.	42
5.4	Estratégia de validação	42
5.5	MetaTrader	43
5.5.1	Coleta e Formação dos Dados.	44
5.6	Base de dados	45
5.6.1	Normalização dos dados	45

5.6.2	Weka . . . . .	50
<b>6</b>	<b>Definição e construção do classificador.</b> . . . . .	<b>51</b>
6.1	Classificadores . . . . .	52
6.1.1	Especificação da técnica . . . . .	53
6.1.2	Efeito Comparativo . . . . .	54
6.2	Abordagem pragmática e formal. . . . .	55
6.2.1	Formal . . . . .	55
6.2.2	Pragmática . . . . .	57
6.2.3	Hibrida . . . . .	59
<b>7</b>	<b>Experimentos, Testes e Resultados</b> . . . . .	<b>60</b>
7.1	Harmonização dos datasets . . . . .	60
7.2	Treinamento e Teste . . . . .	62
7.3	Análise dos Resultados . . . . .	64
7.3.1	Análise do experimento formal . . . . .	64
7.3.1.1	Concordância. . . . .	66
7.3.1.2	Previsões . . . . .	67
7.3.2	Análise do experimento pragmático . . . . .	67
7.3.3	Concordância . . . . .	69
7.3.3.1	Previsões . . . . .	70
7.4	Qualidade das previsões . . . . .	71
7.5	Análise estatística da Abordagem Hibrida. . . . .	73
<b>8</b>	<b>Conclusões</b> . . . . .	<b>75</b>
8.1	Trabalhos Futuros . . . . .	76
	<b>Referências</b> . . . . .	<b>77</b>

# 1 Introdução

A previsão é um termo utilizado desde os primórdios da sociedade, tentar prever os acontecimentos futuros, tais como previsão do tempo, eleições, jogos de azar, dentre outros apresenta-se como uma ambição sociedade moderna.

As técnicas de previsão buscam auxiliar o processo de tomada de decisão entre os responsáveis por desempenhar uma determinada atividade, em especial a previsão de séries temporais. O conceito de séries temporais expressa-se como diversas informações de forma agrupada em uma variável disposta de maneira aleatória dentro de um intervalo temporal específico (BALLINI *et al.*, 2000). As séries temporais são um campo do conhecimento que demonstram-se intrinsecamente ligado as técnicas preditivas, com finalidade de resolver as incertezas da sociedade, minimizando os riscos das possíveis tomadas de decisões (BRESSAN, 2004).

A organização de uma série temporal pode ser representada em sua forma discreta, como o número de atendimento em um determinado estabelecimento ou de forma contínua, como, por exemplo, o registro de previsões do tempo. As séries temporais são encontradas em diferentes áreas do conhecimento, tais como na área da saúde (PRADHAN; PRABHAKARAN, 2017), educação (GOTTARDO; KAESTNER; NORONHA, 2012), ciência da computação (BARRETO *et al.*, 2002), economia e negócios (HIRA; DESHPANDE, 2016). Através da análise preditiva de taxas cambiais, pesquisadores podem aprimorar novas pesquisas em relação às predições e aplicá-las em outras áreas do conhecimento, além de atrair indivíduos com objetivos de retornos financeiros. Para realizar a predição de séries temporais é necessário a elaboração da modelagem do fenômeno a ser estudado e posteriormente a especificação do comportamento da série, estabelecendo as estimativas e análises estatísticas com a finalidade de identificar seu comportamento e suas demais relações (LATORRE; CARDOSO *et al.*, 2001).

O ambiente de estudo considerado para análise das séries temporais do presente trabalho é o mercado Forex (*Foreign Exchange Market*), pois, este mercado apresenta séries temporais de característica estocástica. Um ambiente desafiador para milhões de pessoas, como grupos econômicos, agentes, grupos de investidores, governos e particulares que buscam informações econômicas e financeiras para fundamentar suas tomadas de decisões.

As entidades buscam este mercado com o objetivo de lucrar mais e minimizar suas perdas através de estratégias fundamentadas pela econometria, uma área que relaciona a matemática e estatística com o estudo da economia, no qual estas técnicas aliadas à computação pode viabilizar o sucesso nesta área do conhecimento (MARQUES, 2017).

O Mercado Forex é considerado um mercado de câmbio de divisas, em que é possível a negociação de pares de moedas, tais como: Euro, Dólar, Franco Suíço, Libras Esterlinas, dentre outras de moedas que são disponibilizadas através de uma corretora. É um mercado mundial, onde são negociados cerca de 5 trilhões de dólares diários devido sua alta volatilidade, sendo considerado um dos mais rentáveis em transações de curto prazo, em comparação a mercados como o da bolsa de valores de *New York* (ROLEDENE et al., 2016).

Os usuários considerados comuns e pequenos investidores realizam operações neste mercado através de *homebrokers*. O *homebroker* é uma plataforma disponibilizada por uma determinada corretora escolhida pelo usuário. Através do *homebroker* ele poderá efetuar suas emissões de ordem de compra ou venda, e também utilizar um dos grandes atrativos deste mercado que são as estratégias automatizadas.

As estratégias automatizadas dispõem a possibilidade de elaboração de algoritmos que possam operar no mercado Forex de forma automática em nome do usuário (ROLEDENE et al., 2016). Sendo assim, é de suma importância que os usuários que tenha interesse em adentrar no mercado Forex, estudem o seu funcionamento e comportamento de suas séries temporais. Busquem driblar as armadilhas dispostas no mercado e construa suas estratégias de forma eficientes que resultem em lucros evidentes.

Ao analisar o comportamento das séries temporais disponibilizadas através do mercado Forex é possível a identificação de comportamentos estocásticos, visto que suas variações temporais independem de uma fórmula matemática. Entretanto, é possível perceber uma espécie de equilíbrio em determinados momentos, no qual esses equilíbrios podem ser associados ao comportamento termodinâmico do mercado.

O comportamento termodinâmico do mercado é pautado nos princípios da termodinâmica, ou seja, o mercado não permanece em um estado de equilíbrio de forma definitiva. O mercado pode está em um estado de equilíbrio, quando as suas propriedades permanecem contínuas ao longo do tempo. Também pode apresentar-se de forma transitória, quando suas propriedades sofrem perturbações por conta de fatores externos, tais como notícias e parâmetros heterogêneos, a título de exemplo as comissões das corretoras (NERY; BASSI et al., 2011).

O processo estocástico de predição proposta pelo modelo de análise de estados e transições é feito através da análise markoviana de primeira ordem (ATTA et al., 2018). A análise markoviana de primeira ordem corresponde a previsão de processos, cuja ocorrência de um estado futuro depende apenas do seu estado prévio (MARQUES; SILVA, 2013).

Os modelos markovianos são utilizados a longa data no que diz respeito a temas relacionados a tomadas de decisão, principalmente por causa da sua análise estacioná-

ria que procura por intermédio de parâmetros desconhecidos o reconhecimento de novos padrões (BRIGGS; SCULPHER, 1998). Baseada na estrutura markoviana de primeira ordem, este trabalho busca a transformação do processo do modelo termodinâmico do mercado em uma análise markoviana de segunda ordem, considerando além do estado prévio, o estado antecessor ao estado prévio, definindo uma correlação entre eles. Assim, a correlação entre os estados que servirão como base para a realização da predição de operações no mercado.

Alinhado com a análise dos estados termodinâmicos no mercado Forex, o grande volume de dados gerados pelas séries temporais torna o processo de avaliação destes dados oneroso e artesanal. O surgimento da grande massa de dados não é uma novidade, com a queda dos custos dos componentes computacionais foi possível o armazenamento de quantidades de dados cada vez maiores, assim como a mineração de dados surge como uma proposta que busca transformar esta grande massa de dados em informações úteis (CAMILO; SILVA, 2009).

As técnicas de mineração de dados abordadas na seção 4.3, serão soluções pautadas em tornar os dados facilmente compreensíveis para tomada de decisões, em especial para ambientes estocásticos, pois, a interpretação correta desde dados pode ocasionar em uma vantagem competitiva de fato (ADHIKARI et al., 2016).

De acordo com o que foi explanado ao longo desta introdução, o problema de pesquisa pode ser traduzido da seguinte forma: A construção de um classificador com base no modelo markoviano de segunda ordem dos estados termodinâmicos permite uma previsão lucrativa do mercado Forex?

Este problema no que lhe concerne é fundamentado através de duas hipóteses que corroboram o referencial teórico do mesmo. A primeira hipótese diz respeito a possibilidade de transformação da análise markoviana dos estados termodinâmicos em forma de primeira ordem serem transformadas em uma análise markoviana de segunda ordem. A segunda hipótese aborda a respeito da viabilidade do trabalho, a união entre a teoria e prática, observaremos se as regularidades encontradas no respectivo modelo podem ser aproveitadas para operar no Forex.

Em contraponto ao problema de pesquisa, o objetivo do trabalho vêm para sanar o problema em questão. Assim o objetivo do presente trabalho consiste na definição e construção de um classificador<sup>1</sup> através da aplicação do método de extração de conhecimento em base de dados por intermédio das técnicas de mineração de dados para a predição de tendências, buscando prever as melhores oportunidade de compra ou venda e identificação de falsos positivos no mercado Forex.

Em síntese, a metodologia abordada na realização da presente pesquisa é dividida

---

<sup>1</sup> O classificador em questão será baseado no modelo dos estados oriundo do desenvolvimento de uma estratégia automatizada.

em três partes, na qual a primeira parte consiste na extração da série temporal do mercado Forex, cujo são detectados seus estados através de algoritmos (Estratégias Automatizadas) com base no modelo dos estados termodinâmicos. Neste algoritmo são definidos os melhores parâmetros a serem considerados como variáveis para o classificador. Após a realização desta extração é feita uma análise da base de dados histórica para a construção de uma estratégia de anotação, onde são definidos os critérios de saída a ser interpretado pelos algoritmos de classificação.

A segunda parte é pautada na construção de um arquivo que seja interpretável pelos algoritmos de classificação, onde é testado as estratégias de treinamento mais adequada, buscando ampliar a máxima cobertura de um classificador. Sendo o critério de escolha feito pelo classificador mais otimizado.

Por fim, almeja-se com o presente trabalho, uma contribuição acadêmico-científica que colabore no desenvolvimento de um modelo preditivo eficiente para o mercado Forex, buscando aqui difundir experiências a respeito de ganhos financeiros e mitigação de perdas. Não obstante, o trabalho não está restrito apenas aos usuários e pesquisadores do mercado Forex, e sim, serve como apoio para pesquisadores de séries temporais e mineração de dados.

## 2 Mercado Forex

O mercado Forex é um mercado de câmbio de divisas internacional, este mercado é pautado na negociação de moedas. As negociações realizadas no Forex são estabelecidas através das taxas de câmbio entre diversas moedas do sistema financeiro internacional (ROSSI, 2010). A título de ilustração, são negociados pares de moedas tais como:

Figura 1: Exemplos de moedas negociadas no Forex

Ativo	Bid (preço de co...	Ask (preço de ve...
↓ AUDJPYk	82.528	82.554
↓ AUDUSDk	0.75409	0.75492
↑ EURAUDk	1.54071	1.54111
↓ EURGBPk	0.87343	0.87368
↑ EURJPYk	127.160	127.209
↑ EURNZDk	1.67467	1.67516
↑ EURUSDk	1.16242	1.16283
↓ GBPAUDk	1.76341	1.76419
↑ GBPJPYk	145.576	145.611

Fonte: Elaborado pelo autor.

O Forex é considerado o mercado de maior importância no âmbito global, dado o seu grande volume de operações. São negociados cerca de 5 trilhões de dólares diariamente (ROLEDENE et al., 2016), uma quantidade superior ao mercado de ações e balcão no qual são negociados 6,8 bilhões de reais diariamente (VALOR, 2016). Participam desde mercado investidores diversos, tais como governos, bancos e com o advento da internet foi possível a participação de investidores individuais neste mercado.

O seu funcionamento é de maneira análoga ao mercado de ações. No mercado de ações é necessário que o usuário abra uma conta em uma corretora específica e realize a transferência de dinheiro para a conta desta corretora. Em relação às operações o usuário escolhe a ação que deseja comprar e emite uma ordem de compra no HomeBroker e por fim a BMEFBovespa executa a ordem (INFOMONEY, 2014).

Para ter acesso ao mercado Forex, os usuários também tem que estar associados a uma corretora bem como no mercado de ações, entretanto, no Brasil o mercado Forex ainda não é regulamentado. Contudo, a participação de usuários brasileiros não é proibido, os usuários brasileiros podem associar-se as corretoras estrangeiras para realizar suas operações.

O mercado também é dotado de uma autonomia, pois, detém de uma disponibilidade de 24 horas diárias, durante os 5 dias da semana (de segunda a sexta). Isto possibilita que as transações sejam feitas de forma descentralizada, já que, não existe nenhum órgão

responsável por ser o centro de todas as transações, como é feito no mercado de ações (BORTOLI, 2015).

Silva e Filho (2018) descreve conceitos básicos do mercado Forex, uteis para iniciantes neste mercado.

1. *Day-trade*: É a forma de negociação de modo que seja possível ganhar com a oscilação do preço ao longo do dia
2. *Ask*: É o preço de venda por uma determinada moeda.
3. *Bid*: É o preço de compra por uma determinada moeda.
4. *Spread*: É a diferença entre o Ask e Bid, ou seja, a diferença <sup>1</sup> entre os preços de compra ou venda.
5. *Point*: É o valor fracionário da taxa entre as moedas, aqui que o usuário irá usar como base seus ganhos e perdas.
6. *Buy Order*: É a emissão que efetua a compra de uma moeda, quando o valor aumenta o investidor ganha.
7. *Sell Order*: É a emissão que efetua a venda de uma moeda, quando o valor diminui o investidor ganha.
8. *Hedge*: É uma estratégia de barreira, ou seja, é realizado uma ordem simétrica a uma ordem corrente que tem como objetivo manter a perda constante com a variação do preço.
9. *Take Profit*: É um limite de ganho posto pelo usuário e quando a ordem atinge este valor ela é executada.
10. *Stop Loss*: É um limite de perda posto pelo usuário, onde ele decide qual será a maior perda aceita.

## 2.1 Características

Segundo Rossi (2010) o Forex têm como enfoque cinco características clássicas. A predominância do mercado de balcão, alto grau de liquidez, baixo grau de regulação, alta alavancagem e a ausência de viés altista ou baixista.

Como dito anteriormente, não há a existência de uma entidade centralizadora neste mercado, ou seja, a maioria das suas operações são realizadas no mercado de balcões. Isto

<sup>1</sup> Uma corretora também estipula valores que são acrescidos a esta diferença, boas corretoras possuem valores baixos de spreads

resulta em um baixo grau de regulamentação, já que não existe um registro único de todos os operadores e operações realizadas.

Graças a seu funcionamento 24 horas o Forex propicia a seus usuários um alto grau de liquidez, pois, as suas transações podem ser liquidadas a qualquer momento através das plataformas virtuais. O sistema de alavancagem oferecido pelas corretoras são consideradas margens altas, pois, são oferecidas alavancagens que chegam a propiciar ganhos de cem por cento ou uma perda de igual valor.

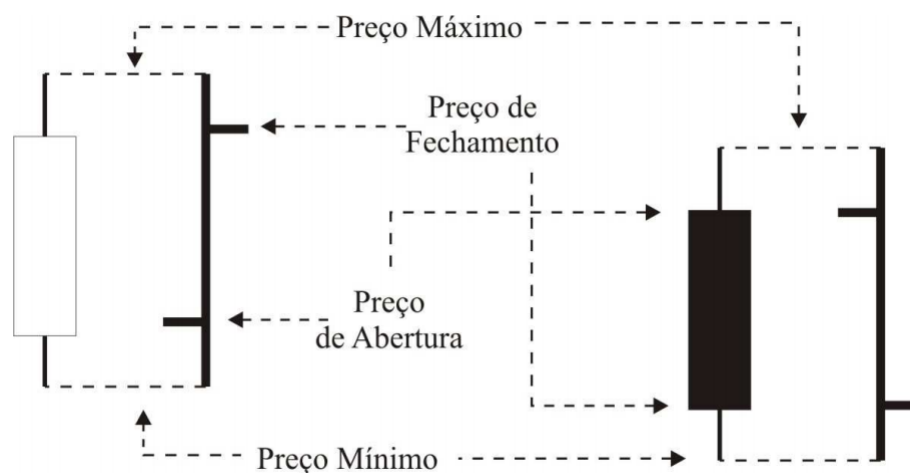
Por fim a sua última característica, pois, de forma distinta ao mercado de ações, que é considerado as ações em alta (com maior valor) ou em baixa (com menor valor), no Forex a compra de uma moeda resulta na venda da outra e vice-versa. Sendo assim, quando uma moeda valoriza-se, automaticamente a outra moeda desvaloriza-se.

## 2.2 Séries temporais do Forex

O mercado forex detém de séries temporais financeiras. Uma série financeira justifica-se através de características de não linearidade e estocasticidade. Bem como possui componentes de sazonalidades e tendências, dotadas de comportamentos pseudo-cíclicos. Convergindo para uma possível não estacionariedade, o que torna difícil a sua previsão (BORTOLI, 2015).

As séries temporais do mercado Forex são representadas através de barras ou velas (*candlesticks*). O tempo de observação pode ser escolhido pelo usuário, ele pode escolher tempos de 1 minuto, 5 minutos, 30 minutos, 1 hora, 4 horas, 1 dia, 1 semana e 1 mês. Através de cada *candlestick* é possível observar cinco grandezas inerentes, tais como o preço de abertura e fechamento do período, preços máximos, mínimos e o volume de negociações (SILVA; FILHO, 2018). Silva e Filho (2018)

Figura 2: Representação dos candlesticks do Forex



Fonte: (SILVA; FILHO, 2018).

## 3 Modelagem de Séries Temporais

O presente capítulo realiza uma investigação em relação à modelagem de séries temporais, em especial as séries temporais financeiras. Uma série temporal financeira é definida através dos ativos decorridos sobre um intervalo de tempo (KEIEL; BENDER, 2018). Sendo assim, ao realizar inferências no que concerne a análise de séries temporais, faz-se necessário à construção de modelos que possam descrever seu comportamento, buscando assim estimar suas possíveis relações causais observando quais são os fatores viáveis que influenciam o seu comportamento estrutural (KOSHIYAMA et al., 2018).

Condizente com o que foi dito anteriormente, a construção de um modelo é de suma importância para a realização de análise de séries temporais, pois, através deles é possível entender o comportamento dinâmico da série temporal, além de agregar novos conhecimentos. Um exemplo disto é a capacidade de previsão de valores futuros de uma série temporal (KEIEL; BENDER, 2018). Por conseguinte, este trabalho aborda uma metodologia para a modelagem de séries temporais baseado em comportamentos termodinâmicos que se relacionam com as características das séries temporais definido através do trabalho de Atta et al. (2018) propondo uma idiosincrasia para as séries temporais oriundas do mercado Forex.

### 3.1 Séries Temporais

De acordo com Latorre, Cardoso et al. (2001) uma série temporal é a representação de dados de forma histórica, ou seja, uma sequência de elementos extraídos em espaços temporais regulares ao longo de um período específico. A importância de uma série temporal para o desenvolvimento de pesquisas é justificada, pois, o reconhecimento de seus padrões desenrolam-se de acordo com os ocorridos contidos em um determinado período de tempo.

As observações feitas através de uma série temporal é caracterizada de forma discreta, ou seja, uma série é considerada discreta quando a análise de seus valores são feitos por meio de pontos temporais específicos, a título de ilustração, o número semestral de formando da Universidade do Estado da Bahia (CARVALHO, 2017). Literaturas, tais como Carvalho (2017), Morettin e Tolo (2006) trazem definições de séries temporais contínuas, entretanto, quando trazemos para o campo das observações para análise e processamento, os dados são convertidos para uma forma discreta.

Coadunante com estas características, as séries temporais de mercados financeiri-

ros são constatadas através de marcações temporais  $T_1, T_2, \dots, T_n$ <sup>1</sup> sendo denotadas por  $X(T_1), X(T_2), \dots, X(T_n)$ <sup>2</sup> isto corresponde a série temporal ter  $n$ <sup>3</sup> valores sucessivos disponíveis para análise. Os dados utilizados para análise de séries temporais financeiras são coletadas de formas discretas, tal que o  $X$  representa a variável aleatória a ser observada e a variável  $T$  representa o instante de tempo atrelado a variável aleatória (OLIVEIRA, 2007).

### 3.1.1 Características e formação.

Além das características previamente apresentadas, as séries temporais também são consideradas univariadas, quando sua observação é feita por uma única variável ao longo do tempo na forma de  $a_0, a_1, \dots, a_{n-1}$  considerando o intervalo de tempo  $\tau \in [0, n - 1]$  ou multivariadas, quando uma série temporal quando sua observação é feita com base em  $a$ , porém, são analisadas  $b$  variáveis a cada instante de tempo, na forma de  $x_{1\tau}, x_{2\tau}, x_{b\tau}$  com  $\tau \in [0, n - 1]$  (ISHII; RIOS; MELLO, 2011).

Sua compreensão é dada de forma determinística, quando é apresentado um comportamento repetido em escalas iguais ou diferentes. Esta característica pode ser descrita através de um modelo matemático buscando descrever o valor exato da série em qualquer instante de tempo (ISHII; RIOS; MELLO, 2011) and (BORTOLI, 2015). Em contrapartida, elas também são definidas como estocástica, cuja as suas observações futuras podem depender de momentos anteriores, porém, pode sofrer *random effects*. Isto significa que seus valores futuros são representados através de probabilidade.

Intrinsecamente as séries temporais oriundas do mercado financeiro apresentam propriedades não observáveis; A tendência, sazonalidade e componente de aleatoriedade (ADHIKARI; AGRAWAL, 2013). A tendência diz respeito as variações do comportamento de uma série temporal, estas variações definem se uma série vai aumentar, diminuir ou estagnar ao longo de um período de tempo (ISHII; RIOS; MELLO, 2011). Podemos concluir com a análise de Adhikari e Agrawal (2013) que a tendência é o próprio movimento da série temporal. Assim como ele traz exemplos para fundamentar essa ideia, tais como:

- (a) Crescimento populacional (tendência crescente).
- (b) Taxas e mortalidades e epidemias (tendências decrescentes).
- (c) Parada cardíaca de um eletrocardiograma (tendência estagnada).

A sazonalidade traz a ideia que o comportamento de uma série temporal tende a se repetir dentro um intervalo de tempo  $\Delta T$ . A título de ilustração Adhikari e Agrawal

<sup>1</sup> A letra T representa instantes de tempo.

<sup>2</sup> A letra X representa as variáveis aleatórias atreladas aos instantes temporais.

<sup>3</sup> Como dito  $n$  representa valores que são discretizados a depender da série temporal em questão.

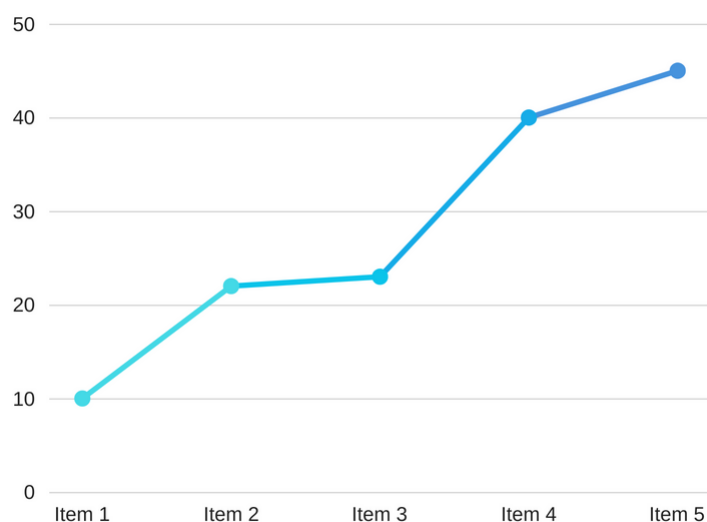
(2013) nos traz que as flutuações sazonais influenciam as séries temporais através de condições climáticas e hábitos culturais. Também nos é apresentado o componente ciclo, considerado uma que extensão desta sazonalidade, cuja diferença é pautada no período de tempo analisado. Por fim o componente de aleatoriedade que remete a ideia de influências imprevisíveis, que não apresentam nenhuma regularidade ou padrões específicos. Tais como guerras, greves, revoluções (ADHIKARI; AGRAWAL, 2013).

### 3.1.2 Processo estocástico e estacionariedade

Alencar (2007) Define o processo estocástico como uma família de variáveis aleatórias<sup>4</sup> dispostas sobre um espaço de probabilidade, classificadas através de um conjunto de parâmetros  $T$ , regida pela seguinte fórmula:  $Z = Z(t)$ ,  $t \in T$ . Não obstante, Ishii, Rios e Mello (2011) elucida que o processo estocástico diverge dos processos determinísticos, pois, os valores futuros de uma série temporal não são determinados por uma função matemática e sim através de distribuições probabilísticas. É definido o processo estocástico através de uma sequência de variáveis aleatórias dispostas em um conjunto  $\Omega$ <sup>5</sup>, definido como um espaço de estado.

São demonstrado graficamente o processo determinístico na imagem 3 e estocástico na imagem 4.

Figura 3: Exemplo de uma série temporal determinística genérica.



Fonte: Elaborado pelo autor.

Como é possível observar na figura 3 são dispostos valores para uma série temporal

<sup>4</sup> Variável aleatória: Uma variável é considerada aleatória ( $Z$ ) é definida como uma função mensurável de um espaço de probabilidade para um espaço mensurável (ALENCAR, 2007)

<sup>5</sup> O  $\Omega$  é definido como um espaço amostral, um conjunto não vazio de todos os resultados possíveis de um experimento (ALENCAR, 2007)

determinística qualquer, ou seja, os seus itens podem ser qualquer elemento relacionado a um tema. Exemplo: número de casas, alunos, dentre outros. Os valores desta imagem 3 são definidos por uma equação de primeiro grau genérica, assim é possível prever seus valores futuros com total exatidão. Bem como na figura 4 é apresentado um gráfico oriundo do Forex, um corte de tempo de 13 de Maio de 2018 até 30 de Maio de 2018. Como é possível observar na figura 4 não há a existência de um gráfico compatível com o nosso sistema de funções matemáticas <sup>6</sup>, pois, sua predição é feita através de dados do passado e os valores futuros são evidenciados através de distribuições probabilísticas (AVENT, 1999).

Figura 4: Exemplo série temporal estocástica do Forex.



Fonte: Elaborado pelo autor.

Os processos estocásticos também apresentam comportamentos ocultos, dentre eles a estacionariedade. Adhikari e Agrawal (2013) traz a luz que a estacionariedade de uma série temporal manifesta-se através do processo estocástico, podendo esta ser visualizada mediante uma forma de equilíbrio estável. Assim como Bortoli (2015) nos apresenta que uma série estacionária traz valores estatísticos, como média e variância inalteráveis ao longo de um intervalo de tempo.

A importância da análise da estacionariedade para o presente trabalho apresenta-se na necessidade da simplificação do processo estocástico, buscando a projeção de um modelo adequado para a predição de séries temporais.

### 3.1.3 Aplicações das séries temporais e importância

O uso da modelagem das séries temporais são provenientes em muitas áreas de estudos, tais como economia, finanças, oceanografia, meteorologia, engenharia, medicina e áreas afins (MORETTIN; TOLOI, 2006). Logo, a depender da natureza ou campo do conhecimento, pode haver diversas aplicações relacionadas as séries temporais. Entretanto, as séries temporais convergem-se no que diz respeito a sua visualização, onde

<sup>6</sup> Os sistemas de funções matemáticas são compostas pelas seguintes funções: polinomial 1 grau, polinomial 2 grau, polinomial 3 grau, exponencial, trigonométrica, modular e logarítmica. (IEZZI; MURAKAMI, 1993)

elas são representadas através de um grafo plotado em relação ao tempo correspondente (ADHIKARI; AGRAWAL, 2013).

A análise de uma série temporal quaisquer, pode trazer *insights* científicos como a investigação do seu mecanismo gerador, previsão de valores; descrição de seu comportamento e análise de sua periodicidade (MORETTIN; TOLOI, 2006).

O presente trabalho busca a observação e predição das séries temporais oriundas do Forex, que detém das características e qualidades citadas anteriormente, entretanto, como visto, existem diversas aplicações de séries temporais em inúmeras áreas do conhecimento, então o presente trabalho poderá contribuir com outros trabalhos não ficando restrito as séries temporais do Forex.

## 3.2 ARIMA

Segundo Koshiyama et al. (2018) o Auto-repressivo Integrado de Média Móvel (ARIMA) é uma família de modelos proposto por Box e Jenkins em 1976, este modelo é pautado em auxiliar a compreensão do comportamento de séries temporais financeiras. Este modelo representa a análise de séries temporais através de modelos autorregressivos integrados e de médias móveis. Além de ser responsável por descrever series temporais não estacionárias.

Esta família de modelo está fundamentada através de uma metodologia que consistem em quatro passos pré definidos: Identificação, Estimação, Checagem e Previsão (KOSHIYAMA et al., 2018). Na identificação é feito o reconhecimento das subclasses para relaciona-las a uma classe geral previamente estabelecida, ou seja, uma classe geral do modelo é apreciada para análise. Após isto são estimados os principais parâmetros com o objetivo de minimizar algum critério de desempenho buscando realizar a identificação do modelo proposto, baseando-se em autocorrelações, dentre outros critérios. Posteriormente após o estabelecimento dos modelos estimados, eles são submetidos a critério de avaliações para que possa ser gerada as saídas adequadas à série e por fim a previsão que consiste em testar os diferentes modelos propostos pelo ARIMA (KEIEL; BENDER, 2018; MORETTIN; TOLOI, 1981).

De acordo com Carvalho (2017) este modelo analisa as características de uma série temporal estacionária partindo do pressuposto que ela detém de uma linearidade seguindo uma distribuição estatística específica. O presente modelo tem uma notoriedade no âmbito de análise de séries temporais, notoriedade justificada, pois, este modelo é versátil para apresentar diversas séries temporais de uma maneira simples.

### 3.3 Redes Neurais

As redes neurais são uma espécie de representação computacional de neurônio humano em forma de nó<sup>7</sup>. Não obstante, [Bortoli \(2015\)](#) define uma rede neural como um modelo não linear, auto-adaptável capaz de generalizar funções matemáticas por intermédio de treinamento e teste utilizando uma base de dados, sendo especialmente utilizadas na predição e modelagem de série temporais.

Já [Carvalho \(2017\)](#) justifica o uso de modelagem de séries temporais utilizando redes neurais artificiais (RNA) por conta que as RNA's buscam solucionar as limitações de modelos lineares, em contraponto a modelarem ARIMA, pois, o processo de modelagem de uma RNA é feita de forma adaptativa e não linear. Isto é viável, pois, uma RNA é capaz de aprender novas tarefas, realizar generalizações e novas descobertas através de entradas.

As utilidades das redes neurais artificiais não se limita à modelagem de séries temporais, também é possível considerar as RNA's como um conjunto de soluções com o objetivo de resolver os problemas de inteligência artificial. Uma forma de dar inteligência a um computador para que ele possa imitar de forma inteligente e autônoma o comportamento humano, apresentando comportamentos imprevisíveis e auto-organizáveis ([BARRETO, 2002](#)).

### 3.4 Outros

Graças a versatilidade e estocasticidade das séries temporais é factível construir modelos proprietários para modelá-la. Evidentemente a construção de modelos proprietários devem respeitar as propriedades previamente citadas na seção 3.1.1 para que haja uma coerência entre o modelo proposto e as propriedades das séries temporais dado o ambiente a ser estudado.

O estudo realizado por [Santos \(2016\)](#) propôs um modelo matemático que serviu como base para uma estratégia de *hedging* para o Forex. A estratégia de *hedging* é pautada na realização da abertura de uma ordem contrária em relação à ordem que se queira compensar. A viabilidade da construção desse modelo foi comprovada através de definições de limites, variáveis relacionadas à abertura e fechamento das ordens e elaboração de relações por intermédio de equações matemáticas.

[Morettin e Tolo \(1981\)](#) analisa séries temporais através do modelo de regressão harmônicas. O presente modelo executa uma aproximação dos valores da série em função do tempo por intermédio da combinação linear de harmônicos, apropriando-se de com-

---

<sup>7</sup> Um nó na computação é a representação de um campo, registro, comumente utilizado para representar árvores e grafos.

ponentes senoidais, ou seja, a ideia aqui é empregar somas trigonométricas, isto é somas de senos e co-senos harmonicamente relacionados buscando descrever fenômenos periódicos (MORETTIN; TOLOI, 1981; AGUIRRE, 2013). Sendo assim os coeficientes das séries temporais são interpretados como as transformadas de Fourier. O presente modelo é utilizado como um auxílio em séries temporais no campo da frequência, realizando uma análise harmônica por meio de combinação de funções trigonométricas.

No trabalho de Oliveira e Favero (2002) é abordado a modelagem de séries temporais através de transformadas de ondaletas (*wavelets*). É explicitado que a *wavelet* é capaz de oferecer informações de tempo e frequência extraídas de uma série temporal de forma simultânea, sendo assim é possível construir uma frequência-tempo da série. Sendo este um modelo adequado para o mapear as mudanças de propriedades em séries não estacionárias

## 3.5 Abordagem proposta no trabalho

A abordagem adotada neste trabalho combina o modelo termodinâmico que considera estados diferentes entre si com o modelo de cadeias de Markov que permite tratar as transições entre esses estados. As seções seguintes descrevem brevemente estes assuntos.

### 3.5.1 Modelo Termodinâmico dos Estados

Os modelos científicos são apresentados através de representações sistemáticas de uma determinada área fim. Com objetivos claros de simplificação de complexidades e reprodução da essência do comportamento do sistema alvo (CARVALHO, 2017). O presente trabalho busca elucidar o modelo dos estados termodinâmicos encontrados nas séries temporais do mercado Forex.

Por intermédio de características como tendência e sazonalidade é possível perceber consonâncias com os princípios básicos da termodinâmica. As conformidades apresentadas justificam-se nos princípios termodinâmicos, pois, a considerar que todos os sistemas são dinâmicos, isto significa que estes não permanecem em um estado de equilíbrio em demasia. Logo, é possível que as séries temporais oriundas do mercado Forex apresentem-se de forma equilibrada, ou em desequilíbrio. Analogamente aos princípios termodinâmicos o possível desequilíbrio, demonstra o aumento, diminuição ou estagnação de sua tendência dentro de um intervalo de tempo.

Sendo assim é possível a classificação de estados básicos voltado a estagnação da tendência da série temporal, respeitando os limites de sua sazonalidade e estados transitórios que nortearam em curtos intervalos de tempo o sentido, seja de subida ou descida.

Tabela 1: Estados identificados nas séries temporais do Forex

<b>Estado</b>	<b>Descrição</b>
Basal	A série permanece oscilando suavemente em volta de um valor constante.
Em subida	A série mostra uma clara tendência para cima.
Em descida	A série mostra uma clara tendência para baixo.
Iniciando subida	Estando em estado basal a série inicia uma tendência de subida.
Iniciando descida	Estando em estado basal a série inicia uma tendência de descida.
Iniciando amortecimento de subida	A inclinação da tendência de subida da série começa a diminuir.
Iniciando amortecimento de descida	A inclinação da tendência de descida da série começa a diminuir.
Amortecendo	A tendência se transforma em oscilações não necessariamente simétricas em volta de um valor de equilíbrio com amplitudes decrescentes no tempo.
Amorteceu tendência	As oscilações são simétricas e estabiliza-se com baixa amplitude.
Reagindo ao amortecimento - subindo	No início do amortecimento ou durante ele a série inicia abruptamente uma tendência de subida.
Reagindo ao amortecimento - descendo	No início do amortecimento ou durante ele a série inicia abruptamente uma tendência de descida.
Reagindo a tendência - subindo	A tendência de descida da série passa abruptamente para tendência de subida.
Reagindo a tendência - descendo	A tendência de subida da série passa abruptamente para tendência de descida.

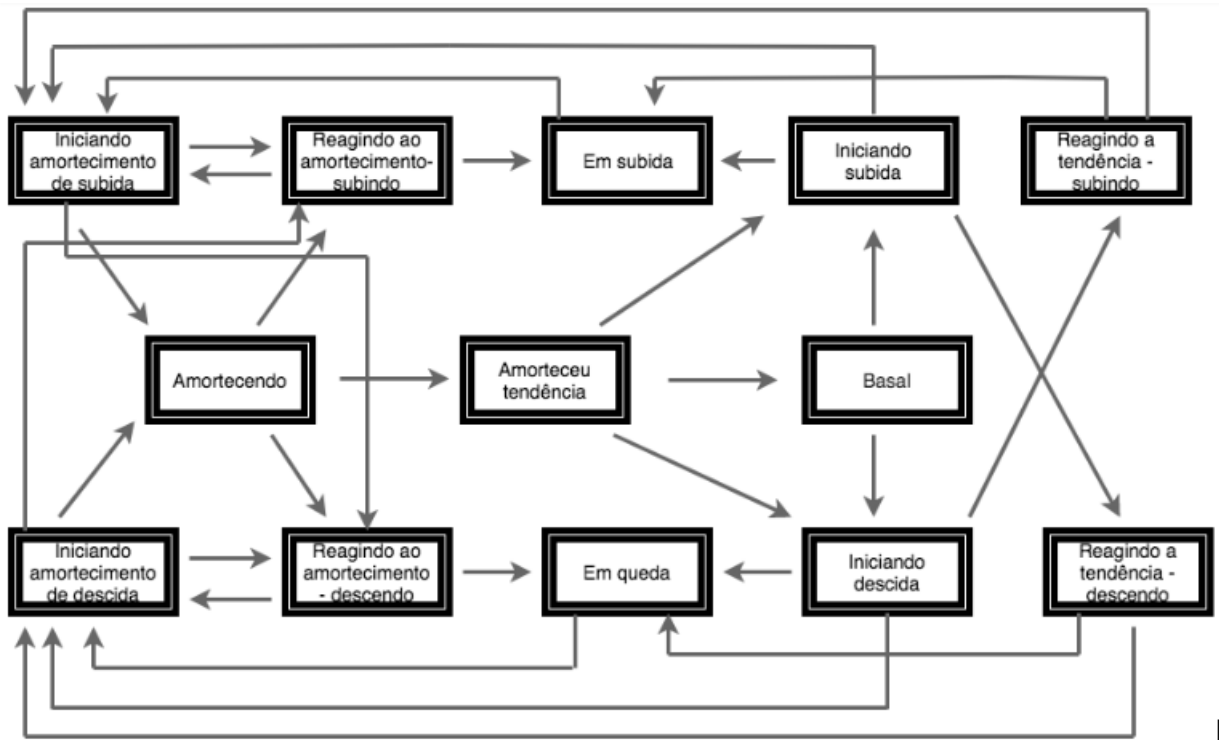
A Tabela 1 ilustra a correlação entre as séries temporais oriunda do Forex e seus respectivos estados.

Bem como na figura 5 é apresentada uma ilustração em relação de como ocorre a comunicação entre os estados.

### 3.5.2 Cadeias de Markov

A análise markoviana surge como solução para problemas de decisões que envolvem incertezas. Os modelos de Markov podem ser utilizados para a realização de análises econômicas, pois, ele possibilita a inclusão de análises temporais (SATO; ZOUAIN, 2010). Sendo assim, os processos markovianos surgem como um caso particular dos processos estocásticos abordados previamente na seção 3.1.2. Na literatura Marques e Silva (2013)

Figura 5: Exemplo dos 13 estados e suas relações.



Fonte: Elaborado pelo autor.

caracteriza a transformação de processo estocástico em markoviano quando o acontecimento de um estado futuro depende estritamente do seu estado prévio.

Um ponto importante para o entendimento da análise markoviana na realização de inferências estatísticas é a caracterização de suas cadeias e classificação de estados. As cadeias de markov são definidas como um processo estatístico cuja probabilidade de mudança de um estado para outro seja constante no tempo e só dependa dos estados, ou seja, esta cadeia representa que, dado  $t_{ab}$  a probabilidade de mudança do estado  $a$  para o estado  $b$  em uma unidade de tempo (SANTOS, 2008).

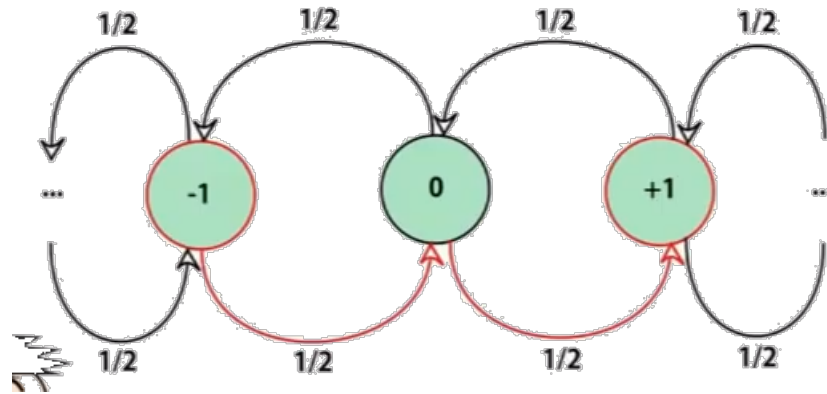
A análise de classificação dos estados de uma cadeia markoviana é verificar se um estado  $x$  é acessível<sup>8</sup>, ou seja, o estado  $x$  é considerado acessível, quando é notado a existência de um estado  $y$  e existe um caminho que leva o processo de um estado para outro. Se os dois estados  $x$  e  $y$  conseguem ser acessíveis entre si, ou seja,  $x$  é acessível a  $y$  e  $y$  é acessível a  $x$ , definimos este estado como comunicado. Probabilisticamente é possível definir que um estado acessível detém da probabilidade de um estado é acessível a outrem, se para algum  $0 \leq n < \infty, \mathbb{P}_{a,b}^n > 0$  (ACTION, 2015; SIGNORETTI, 2013).

A título de ilustração é possível identificar o estado acessível através de um jogo de cara ou coroa, cuja as regras do jogo é ganhar(+1) quando se tira cara na moeda e perder(-1) quando se tira coroa. Assim é possível desenhar a seguinte cadeia correspondente na

<sup>8</sup> Tem autores que nomeia este estado como alcançável.

figura 6, onde os estados representam o ganho ou perda e todos os estados desta cadeia são alcançáveis por todos os outros. É possível visualizar que a probabilidade de mudança<sup>9</sup> de um estado para o outro é de 50% (SIGNORETTI, 2013).

Figura 6: Ilustração do estado acessível através de um exemplo de jogo cara ou coroa



Fonte: (SIGNORETTI, 2013).

Segundo Marques e Silva (2013) os Estados de uma cadeia markoviana é dotado das seguintes características: Ser recorrente, periódico, transiente e absorvente.

Um estado  $x$  é considerado recorrente<sup>10</sup>, quando o próprio estado é capaz de retornar a ele mesmo com probabilidade 1, ou seja, uma vez que o processo passa por um estado qualquer, existe uma chance do estado  $x$  voltar a ele mesmo com probabilidade de 1. É possível notar na figura 6 que o estado 0 consegue retornar a ele mesmo com probabilidade 1<sup>11</sup> (SIGNORETTI, 2013; MARQUES; SILVA, 2013).

Segundo Action (2015) matematicamente um estado recorrente é representado como: se existe um  $n > 0$  tal que  $f_{i,i}^* = 1$  podemos definir na seguinte forma:  $f_{i,i}^n = \mathbb{P}[X_n = i, X_r \neq i \mid X_0 = i]$  com  $r \in \{1, 2, \dots, n-1\}$  então é possível formar:

$$f_{i,i}^* = \sum_{n=1}^{\infty} f_{i,i}^n \quad (3.1)$$

Assim é possível perceber que  $f_{i,i}^n$  significa a probabilidade de começarmos em um estado  $i$  e retornamos para  $i$  em um tempo  $n$ . Em termos de probabilidade temos que um estado é recorrente se e somente se  $f_{i,i}^* = 1$  como dito anteriormente (ACTION, 2015). Por isto é possível dizer que quando se entra neste estado um eventual retorno é assegurado (NOGUEIRA, 2009). Uma inferência interessante ao se trabalhar com análise de estados recorrentes em uma cadeia markoviana é a definição da periodicidade, ou seja, a determinação do tempo de retorno de um estado  $x$  para ele mesmo. Sendo assim, seu

<sup>9</sup> representado pela seta vermelha

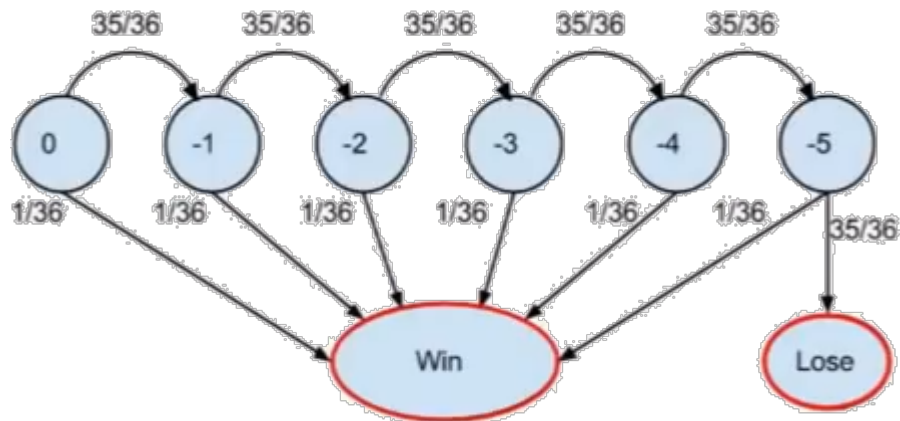
<sup>10</sup> Só é possível um estado ser recorrente se, e somente se, o estado não for transiente.

<sup>11</sup> Basta somar a probabilidade de ida 1/2 e a de volta 1/2.

tempo é marcado em  $t > 0$  e seu retorno é possível em  $t, 2t, 3t, \dots$  passos (MARQUES; SILVA, 2013).

Não obstante, os estados recorrentes também apresentam a características de se transformarem em estados absorventes, um estado é considerado absorvente quando é possível retornar a ele mesmo com uma certeza de transição, ou seja, quando o processo entra neste estado ele nunca mais o deixa-o (MARQUES; SILVA, 2013; NOGUEIRA, 2009). É possível exemplificar o estado transiente como um jogo de roleta em um cassino, cujo jogador tem a possibilidade de escolher um número entre 1 e 36, caso o jogador acerte ele ganha o jogo. A cadeia de markov ilustrada na figura 7 representa o jogo, onde o jogador só pode jogar no máximo até 6 fichas, ganhando ele parara de jogar. Na presente cadeia cada transição representa uma nova aposta e cada estado representa os ganhos no jogo, logo é possível notar com a figura 7 o estado *Win* e *Lose* são considerados estados absorventes, pois, uma vez alcançados nunca serão deixados (SIGNORETTI, 2013).

Figura 7: Estados transientes e recorrentes.



Fonte: (ACTION, 2015).

Em contraponto aos estados recorrentes existem os estados transientes. Um estado  $x$  é considerado transiente quando é possível que ele acalce um outro estado, mas não consiga voltar a ele mesmo com base em outros estados, sendo assim um eventual retorno ao estado  $x$  não está assegurado (MARQUES; SILVA, 2013; NOGUEIRA, 2009). Probabilisticamente para um estado ser transiente a sua probabilidade de retorno tem que ser  $<1$  e sua probabilidade de não retornar ao estado  $x > 0$ , isto garante que o processo pode até não retornar, ou seja, o retorno ao estado pode eventualmente nem ocorrer. Logo o tempo de retorno não é bem definido neste estado (SIGNORETTI, 2013).

Assim Action (2015) nos diz que é possível demonstrar matematicamente que, um estado é considerado recorrente se

$$\sum_{n=1}^{\infty} \mathbb{P}i, i^n = \infty \quad (3.2)$$

e transiente se

$$\sum_{n=1}^{\infty} \mathbb{P}i, i^n < \infty \quad (3.3)$$

Então de acordo com [Action \(2015\)](#) é possível constatar que os estados se equivalem em  $\sum_n f_{i,i}^n = 1$  e  $\sum_n \mathbb{P}_{i,i}^n = \infty$ , nos mostrando a distinção entre  $f_{i,i}^n$  e  $\mathbb{P}_{i,i}^n$ . Assim  $f_{i,i}^n$  nos remete a probabilidade do primeiro retorno de  $i$  em  $n$  passos e  $\sum_n f_{i,i}^n$  sendo a probabilidade do processo retornar eventualmente ao estado  $i$ . Em contrapartida, podemos dizer que  $\mathbb{P}_{i,i}^n$  é referente a probabilidade do evento ocupar o estado  $i$  depois de  $n$  passos. Se o processo retorna ao estado  $i$  no mínimo uma vez, é possível inferir que ele pode retornar outras vezes, assim temos que  $\sum_n \mathbb{P}_{i,i}^n$  é uma soma infinita de números positivos. Portanto, seja  $Q_{ii}^{(N)}$  a probabilidade de que a partir de um estado  $i$ , o processo retorne a este estado  $N$  vezes, então tomamos  $N \rightarrow \infty$  sendo a probabilidade de retorno infinitas vezes ao estado  $i$ . Então temos para um estado recorrente

$$\lim_{N \rightarrow \infty} Q_{ii}^{(N)} = 1 \quad (3.4)$$

e para um estado ser transiente

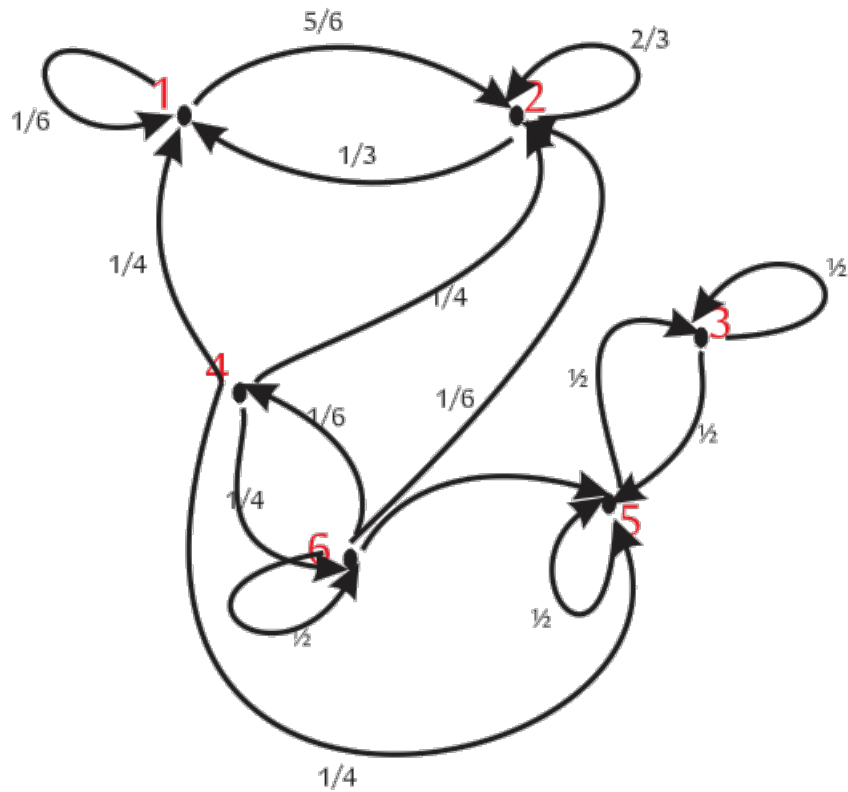
$$\lim_{N \rightarrow \infty} Q_{ii}^{(N)} = 0 \quad (3.5)$$

A título de ilustração figura 8 representa uma diferenciação gráfica de estados transientes e recorrentes. Podemos observar a existência de um modelo  $S$  com 6 estados, ou seja,  $S = \{1, 2, 3, 4, 5, 6\}$  cujos estados estão representados em vermelho, as setas representam as transições entre os estados e as frações acima das setas representam as probabilidades de mudança de um estado para outro. Separando os estados em classes é observado que pelo diagrama as classes  $C_1 = \{1, 2\}$  e  $C_2 = \{3, 5\}$  são recorrentes, pois, a soma de suas probabilidades resultam em 1 e a classe  $C_3 = \{4, 6\}$  é transiente, pois a soma de suas probabilidades estão entre valores  $i > 0$  e  $i < 1$  ([ACTION, 2015](#)).

Neste presente trabalho a série temporal oriunda do Forex será considerada de forma discreta e os eventos serão representados através dos estados explanados na tabela 1. Assim será possível modelar os eventos cíclicos e a dependência temporal da probabilidade no estado atual e antecessor possibilitando uma predição mais precisa do modelo dos estados termodinâmicos.

Com base no que foi apresentado ao longo deste capítulo é possível estabelecer uma relação entre o modelo dos estados termodinâmicos proposto por [Atta et al. \(2018\)](#) e a análise markoviana. Podemos considerar este modelo como sendo uma cadeia markoviana com os estados sendo representados como nós, cuja cada transição representa a ligação entre os nós. Também é possível observar que esta cadeia markoviana formada pelos estados

Figura 8: Estados transientes e recorrentes.



Fonte: (ACTION, 2015).

detém de características, tais como a acessibilidade, já que é possível um estado acessar outem e recorrência, pois, ao *passar* entre os estados é possível que um estado retorne a ele mesmo. A grande incógnita apresentada aqui para a realização de uma previsão assertiva é a periodicidade, estabelecendo a periodicidade dos estados corretamente, será possível prever quando eles irão acontecer.

## 4 Mineração de Dados.

A mineração de dados surge como uma área de pesquisa multidisciplinar que inclui o uso da tecnologia de banco de dados, integrando com as demais áreas da computação, em especial reconhecimento de padrões, sistemas baseados em conhecimento e visualização de dados (AMO, 2004). Seus objetivos claros de automação e análise informações, possibilita a extração de conhecimento útil e desconhecido em meio ao caos, por intermédio de algoritmos que são capazes de extrair modelos e padrões representativos destes. (ALENCAR, 2007). Amo (2004) nos traz um exemplo prático da mineração de dados no meio corporativo: "Uma empresa de cartões de crédito pode saber dos hábitos dos clientes. O que ele costuma consumir, qual seu padrão de gastos, seu grau endividamento". E também nos apresenta as principais razões para a utilização da mineração.

1. O volume de dados são de suma importância para as organizações e por isso não devem ser tratados apenas com técnicas clássicas de análise de dados.
2. Os usuários finais nem sempre possui conhecimentos estatísticos necessários para a realização de uma análise de dados precisa e consistente.
3. Os grandes tráfegos de dados na internet possibilita o acesso a dados de forma exponencial

De acordo com SOUZA (2011) a mineração de dados também nos permite formular relacionamentos sistemáticos utilizando variáveis, buscando à validação dos dados através da aplicação de subconjuntos de dados. Subconjuntos estes que estão ocultos em situações que envolvem grandes conjuntos de dados.

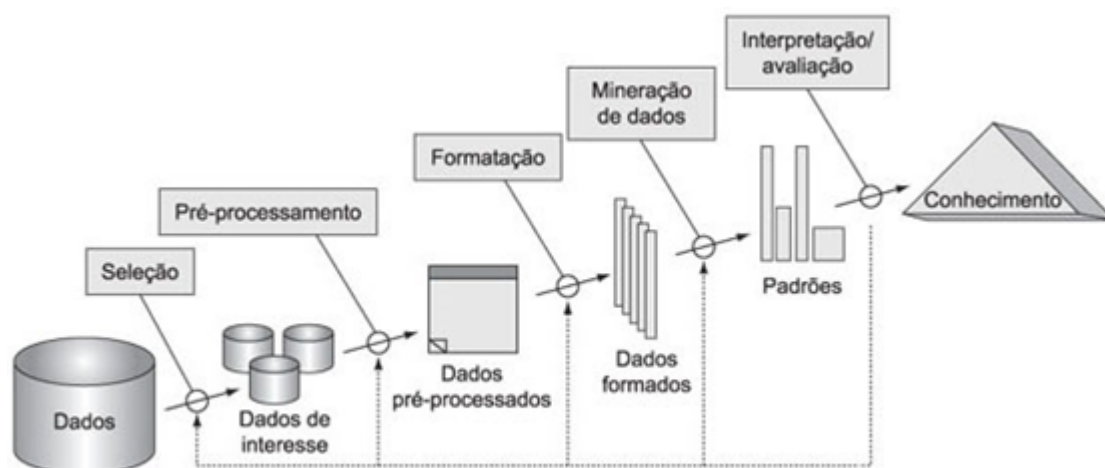
Dentro do contexto da mineração de dados, uma técnica de suma importância é o aprendizado de máquina (*Machine Learning*). Esta sub-área da inteligência artificial, busca a criação de algoritmos que tem por objetivo aprender e analisar de forma autônoma ou supervisionada, características, regras, exemplos, padrões dos ambientes nos quais eles estão submetidos. Logo, ao fornecermos uma quantidade aceitável, de preferência com muitos exemplos, para estes algoritmos é possível que eles realizem inferência e predições, auferindo conclusões das relações(ou subconjuntos) fornecidas, expressando o que de fato importa para as tomadas de decisões (SOUZA, 2011).

## 4.1 Descoberta de Conhecimento em Banco de Dados

Quando falamos de mineração de dados, um dos pontos fortes deste processo independente do seu método é o KDD, cujo ele é pautado nas seguintes etapas ilustradas na figura 9. Amo (2004) descreve a abordagem e etapas no KDD nos seguintes passos.

1. Limpeza dos dados: Nesta etapa busca-se eliminar os ruídos e dados inconsistentes.
2. Integração dos dados: Nesta etapa busca-se combinar diferentes fontes de dados que podem ser combinadas, criando-se um único repositório de dados.
3. Seleção: Busca-se selecionar os atributos que realmente interessam ao usuário <sup>1</sup>.
4. Transformação dos dados: Nesta etapa busca-se normalizar os dados para um formato apropriados, para posteriormente ser aplicado os algoritmos de mineração de dados.
5. Mineração: A etapa essencial do processo, cuja é feita a aplicação das técnicas de mineração de dados com o objetivo da extração de padrões de interesses.
6. Avaliação ou Pós-Processamento: Etapa com o objetivo de identificar os padrões de acordo com os critérios previamente estabelecidos.
7. Visualização dos Resultados> Etapa com o objetivo de aplicar as técnicas de apresentação de conhecimento.

Figura 9: Exemplo do processo KDD.



Fonte: Teofilo (2015)

<sup>1</sup> Exemplo: usuário pode decidir que informações como endereço e telefone não são relevantes para decidir se um cliente é um bom comprador ou não.

## 4.2 Mineração de dados aplicadas as séries temporais

As séries temporais oriundas dos mercados financeiros são consideradas complexas, o que torna difícil sua manipulação. Esta complexidade é justificada através de características, tais como: Alta dimensionalidade, pois, a dimensionalidade da série temporal ta estritamente atrelada ao seu tamanho, ou seja, quanto mais longa uma série mais difícil será sua análise; Grande conjunto de dados, bem como dito anteriormente, a dimensionalidade da série pode acarretar em grande conjunto de dados e existência de series com vários gigabytes; Subjetividade, pois, quem por muitas vezes define o que é certo ou errado é o próprio usuário (ALENCAR, 2007).

## 4.3 Técnicas de Mineração de dados

As técnicas de mineração de dados são pautadas no objetivo de especificar os métodos que nos auxiliem a descobrir quais padrões nos interessam. O presente trabalho abordará as regras de associação. Uma regra de associação diz respeito a relacionar padrões X e Y em forma de implicação, onde consideramos X E Y valores diferentes. Refletindo na busca de um padrão relacionado aos comportamentos (AMO, 2004). A segunda técnica abordada será a de classificação e predição, na qual esta regra é pautada em modelar um conjunto de classes buscando prever os objetos que ainda não foram classificados (AMO, 2004). O modelo construído é baseado em uma análise prévia de um conjunto de amostragem ou dados de treinamento, que irá conter os objetos corretamente classificados. (AMO, 2004). Nesta técnica Também é possível buscar valores ausentes da base de dados.

Uma outra visão acerca das técnicas e metodologias aplicadas para a mineração de dados é explanada por SOUZA (2011) que define as metodologias para a realização da mineração de dados de acordo com o nível de conhecimento do ambiente e/ou situação;

A primeira metodologia é a descoberta não-supervisionada. É recomendado o uso desta metodologia quando não se tem ciência em relação ao objetivo proposto, ou seja, o pesquisador encontra-se em uma situação onde ele não tem domínio nenhum sobre seus dados. Logo, ao encontra-se neste estado, sugere-se a aplicação de algoritmos de mineração de dados que realizem inferências de forma automática.

A segunda metodologia é baseada no teste de hipótese, nesta metodologia o pesquisador detém de algum conhecimento sobre o ambiente/domínio no qual ele está realizando a mineração de dados, logo é definido hipoteses, buscando verificar sua aceitação ou refutação. Por fim a Última metodologia é a modelagem de dados, que é sugerida quando o pesquisador tem total domínio ou um grau de conhecimento elevado em relação à área a ser estudada.

Por outro lado, as técnicas utilizadas para aplicar a mineração de dados são abran-

gentes e adaptativas, podendo serem implementadas por inúmeras ferramentas ou até de forma proprietária. A seguir são explanadas as técnicas para a realização da mineração de dados.

### 4.3.1 Classificação.

Alencar (2007) traz a definição de classificação: Dado um conjunto de  $n$  dados  $X = x_1, x_2, \dots, x_n$ , cujas  $y$  classes são conhecidas como  $Y = 1, 2, \dots, J$ . O objetivo da classificação é o aprendizado de uma função  $f : X \rightarrow Y$ , que mapeie um objeto  $x \in X$  para a sua classe  $y \in Y$ . A presente função será definida como o modelo de classificação.

A classificação é uma das técnicas de mineração de dados que tem como principal propósito prever uma potencial classe fundamentando-se através de uma base de dados histórica. Esta técnica é delineada através de três etapas, que segundo Amo (2004) são elas:

1. Criação do Modelo: Nesta etapa é criado o modelo que será constituído de regras que permitirá a classificação das tuplas no banco de dados através de um número de classes pré-determinados. O modelo é criado por intermédio de um banco de dados de treinamento, que são consideradas as amostras ou exemplos.
2. Etapa de Classificação: É a etapa de execução propriamente dita, aqui é realizado os testes das regras sobre um outro banco de dados, independente do banco de dados de treinamento. A qualidade do modelo será medida através das porcentagens entre as tuplas deste novo banco de dados, verificando se as regras estão conseguindo classificar de forma satisfatória.

De acordo com Alencar (2007) é possível concluir que com este modelo é possível categorizar corretamente tanto os utilizados da etapa de treinamento, quanto os dados jamais vistos antes, oferecendo o benefício de uma boa capacidade de generalização.

### 4.3.2 Associação

SOUZA (2011) descreve a associação como uma técnica de mineração de dados pautada na inferência de padrões através da descrição de características intrínsecas dos dados, ou seja, os padrões são formados por intermédio de regras de implicações ou sub-conjuntos das características, com o principal objetivo de obter-se padrões interessantes de uma forma simples e eficiente.

Sendo assim, na associação é estabelecido a identificação do relacionamento dos atributos mais frequentes dentro um conjunto de dados específicos. As associações são comumente utilizadas nas análises para dispor produtos em um supermercado, a título

de ilustração é possível descrever uma regra do tipo: clientes que compram pão, também compram manteigas. Contudo, é possível descrever um algoritmo de associação da seguinte forma: primeiro realizar um estudo dos itens que aparecem frequentemente, respeitando um limiar mínimo para frequência e posteriormente gerar regras válidas de acordo com a mineração desde conjunto (CAMILO; SILVA, 2009).

### 4.3.3 Agrupamento

O agrupamento é uma técnica fundamentada no estabelecimento de grupos de dados (*clusters*) que estão diretamente relacionados entre si, sendo assim os subconjuntos de dados gerados formam os *clusters* que são definidos por intermédio das similaridades dos atributos do conjunto de dados, ou seja, as similaridades entre os dados são utilizados para definir um conjunto finito de categorias que os defina (SOUZA, 2011).

Outra questão do agrupamento que, ao mesmo tempo, que ele une os dados a partir de suas similaridades, os grupos são formados também por conta das suas diferenças. Já que um grupo é diferente dos demais grupos. É uma boa técnica para o aprendizado não supervisionado, pois, os dados não precisam ser previamente categorizados. A técnica de agrupamento é bem-vinda quando queremos apenas identificar grupos ou comportamentos de dados similares, um exemplo disto é quando ela é usada para segmentar um mercado para um nicho de produto ou em auditoria quando queremos separar comportamentos suspeitos (CAMILO; SILVA, 2009).

## 4.4 Técnica proposta no trabalho

No presente trabalho buscou-se abordar à técnica de mineração de dados fundamentada na classificação, previamente abordada na seção 4.3.1 por conta da diversidade de abordagens são utilizadas para prever um conjunto de dados específico (*datasets*).

(SOUZA, 2011) ilustra os classificadores com abordagens distintas para a realização de tarefas, tais como os classificadores baseados em árvores de decisão, responsáveis por definir um comportamento determinístico, construindo uma árvore com base nas informações deduzidas dos *datasets*. Os classificadores fundamentados na implementação de redes neurais artificiais, que não conhecem a distribuição, realizando uma inferência estatística dos *datasets*. Os classificadores bayesianos que também apresenta uma abordagem determinística, baseando-se na inferência probabilística e por fim os classificados de vizinhança próxima, que tem por objetivo encontrar todos os exemplos dos *datasets* de treinamento semelhante aos atributos dos *datasets* de testes (SOUZA, 2011).

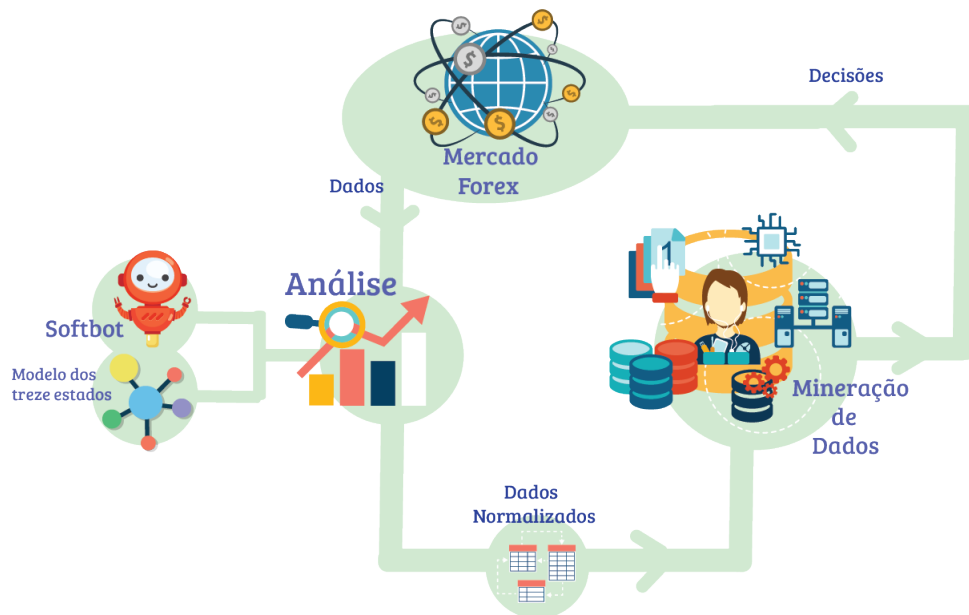
A adoção da classificação justifica-se por uma abordagem que traz elementos como precisão, velocidade, robustez, escalabilidade e interpretabilidade, além de trazer elementos baseados em estatística preditiva para a realização de validações e testes. Também

é possível por intermédio desta técnica à generalização e adaptação de modelos, tanto supervisionados quanto não supervisionados, onde cada classificador traz abordagens diferenciadas dado um determinado conjunto de dados ou variáveis a serem estudadas. (SOUZA, 2011)

## 5 Realização da previsão

O modelo preditivo que se busca no presente trabalho com a classificação será auferido através de um estudo realizado em duas partes. Uma primeira parte que envolve o Mercado Forex, *Expect Advisors* e o MQL4. Não obstante, esta primeira parte conta com a análise da série temporal propriamente dita, a detecção dos estados e estabelecimento de seus parâmetros juntamente com a criação de uma estratégia de anotação, esta estratégia pode ser traduzida com o que se espera baseada nos estudos e experimentos. A segunda parte diz respeito ao envolvimento da mineração de dados, nesta etapa envolve as estratégias de treinamentos e testes para a criação e otimização dos classificadores, o método que se traduz nos algoritmos de fato escolhido e o modelo que é o classificador propriamente dito, ou seja, é a resultante do experimento, o possível classificador que irá dizer se é viável a realização de operações no mercado.

Figura 10: Ilustração do modelo.



Fonte: Elaborado pelo autor adaptado de [Carvalho \(2017\)](#).

Bem como na imagem 10 pode-se verificar que o modelo começa com a realização da extração dos dados do presente objeto de estudo, o Forex, dados estes obtidos de uma plataforma de negociação eletrônica. Com a coleta destes dados históricos, eles serão reorganizados com o intuito de lograr uma base de dados de acordo com as análises feitas com os algoritmos de detecção de estados. Após a realização desta fase de análise é feita a mineração de dados com base nos algoritmos de classificação, onde os algoritmos irão receber e processar as informações contidas na base de dados e produzir o conjunto de dados de saída. Estes dados em saídas estarão coadunantes com os resultados previstos nas

estratégias de anotação, ou seja, esses dados serão as operações de abertura e fechamento de ordens de compra ou venda do Forex.

## 5.1 Metodologia

O desenvolvimento da metodologia está fundamentada por [Hinojosa \(2010\)](#), que define uma metodologia científica para desenvolvimento de projetos. Sendo assim, a metodologia empregada neste trabalho conta com quatro fases principais: Modelagem, extração de dados, aplicação das técnicas de mineração de dados e validação. Mostraremos durante o percurso desta seção cada etapa do trabalho.

A fase de modelagem tem como objetivo a implementação de um softbot <sup>1</sup> capaz de reproduzir os estados termodinâmicos que foram explanados no capítulo 3.5.1. Em seguida será feita a aplicação deste softbot no mercado Forex, implicando na realização da coleta de dados. Esta primeira parte envolve as séries temporais, a detecção dos estados com o estabelecimento de parâmetros e criação dos métodos preditivos, ou seja, esta primeira parte está ligado diretamente a fase estrutural da pesquisa.

Realizada a coleta de dados, é feita a extração de dados com base no método de descoberta de conhecimento, previamente abordado no capítulo 4.1. Assim é possível normalizar a base de dados para que sirva de entrada para os algoritmos de mineração de dados. Seguindo o método de descoberta do conhecimento é realizada a mineração de dados de fato, onde são processados os dados e retirado as análises e resultados.

## 5.2 Modelo

A fim de alcançar os objetivos propostos no presente trabalho, será seguido um roteiro com base na figura 14. Sendo apresentado de forma prática o modelo implementado, ou seja, a demonstração do processo desde o seu início com revisitação da revisão sistemática e literatura, até a análise e comparação de resultados.

O primeiro estágio do trabalho teve o intuito de coletar a série temporal do Forex em forma de dados históricos. Tendo como fonte a criação do modelo dos estados termodinâmicos. Neste estágio a construção do modelo foi feita com a linguagem de programação MQL4.

O segundo estágio teve como foco a definição e aplicação dos algoritmos de classificação. Os Algoritmos de classificação escolhidos foram J48, NaiveBayes, MultilayerPerceptron, KStar e ZeroR.

---

<sup>1</sup> É um agente inteligente capaz de operar autonomamente e efetuar tarefas singulares sem a direta supervisão humana. ([NÉRI, 2005](#))

Figura 11: Ilustração do roteiro metodológico.



Fonte: Elaborado pelo autor adaptado de [Carvalho \(2017\)](#).

O terceiro estágio foi o de verificação, após a realização dos testes foi atestado se com a identificação do método da família de padrões de segunda ordem atrelada a abordagem baseada nas amplitudes dos pips <sup>2</sup>.

Por fim no último estágio buscou-se analisar os resultados com base no classificador proposto, realizando previsões e atestando a adequação da mineração de dados quanto ao objetivo de prever as séries temporais do mercado.

### 5.3 Revisão Sistemática e Literatura.

A revisão sistemática foi a primeira etapa para a elaboração deste trabalho, pois, teve como principal objetivo o levantamento da eficácia/eficiência da aplicação das técnicas de mineração de dados aplicadas nas séries temporais no mercado forex. Assim foi possível verificar a viabilidade da pesquisa e embasar toda fundamentação teórica.

### 5.4 Estratégia de validação

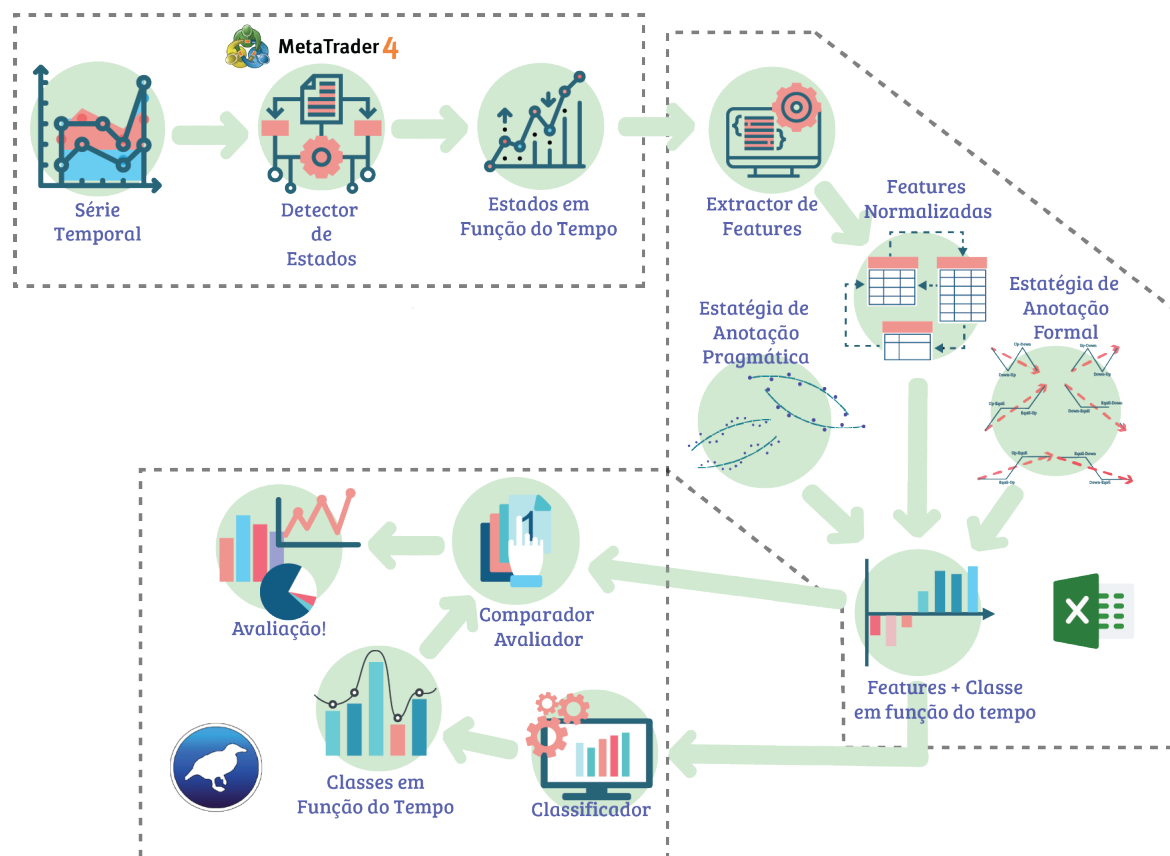
De acordo com o que foi apresentado na figura 14 é importante salientar que além dos cinco estágios anteriormente citados, existe sub-estágios que estão adentrados nas etapas de extração de dados, construção do classificador e validação. Onde esses estágios compõem de fato a estratégia de validação para o presente trabalho. A estratégia de validação pode ser esquematizada através da figura 12.

Dado ao grande volume de dados gerados pelas séries temporais do Forex, pode vir a ser inviável a análise dessas séries caso um pesquisador queira explorá-las e modela-

<sup>2</sup> Pip é o valor unitário da última casa decimal da cotação a ser negociada, essa nomenclatura é muito utilizada para a mensuração de valores no Forex.

las. Por conta disto, foi criada uma estratégia de validação. A estratégia de validação em questão vem para sanar os objetivos principais propostos pelo trabalho, bem como, auxiliar pesquisadores, especialistas e interessados nessa área de estudo.

Figura 12: Ilustração da estratégia de validação.



Fonte: Elaborado pelo autor.

O objetivo desta estratégia de validação é a avaliação das métricas e resultados obtidos por intermédio dos classificadores propostos. Porém, para obter estes resultados, faz-se necessário o cumprimento deste plano. No presente trabalho, como é possível observar na figura 12, foi utilizada uma tríade tecnológica; O *MetaTrader4* responsável por analisar, extrair e modelar a série temporal. O Excel, encarregado de manter e manipular a base de dados. Por fim o Weka, que ficou encarregado pela construção, geração de resultados e avaliação.

## 5.5 MetaTrader

A primeira etapa do plano de validação é a análise e extração das séries temporais do Forex, como dito anteriormente, esta parte foi encarregada pelo *MetaTrader*. Primeiro para a extração da série temporal, é necessária a associação a uma corretora, na qual a corretora em questão disponibiliza um *homerbroker*, sendo possível a construção de *Expects Advisors*. No presente trabalho a série temporal é processada por um algoritmo

definido como um detector de estados. O detector de estados retorna a série temporal modelada em função do tempo. Este processo é detalhado por meio da coleta e formação da base de dados abordado na sub-seção seguinte.

Figura 13: Ilustração da etapa do MetaTrader.



Fonte: Elaborado pelo autor.

### 5.5.1 Coleta e Formação dos Dados.

A coleta das séries temporais do projeto é oriunda do Forex, com base no capítulo 2, o mercado Forex possui uma plataforma chamada de *homebroker*, nesta plataforma os negociadores realizam suas operações e desenvolvem estratégias automatizadas. Sendo assim é possível a extração das séries temporais, através de desenvolvimento de algoritmos em MQL4. Em especial, neste trabalho, a extração da série temporal deu-se através *softbot* implementado na linguagem MQL4 que representa o modelo dos estados termodinâmicos que já fora abordado previamente no capítulo 3.5.1.

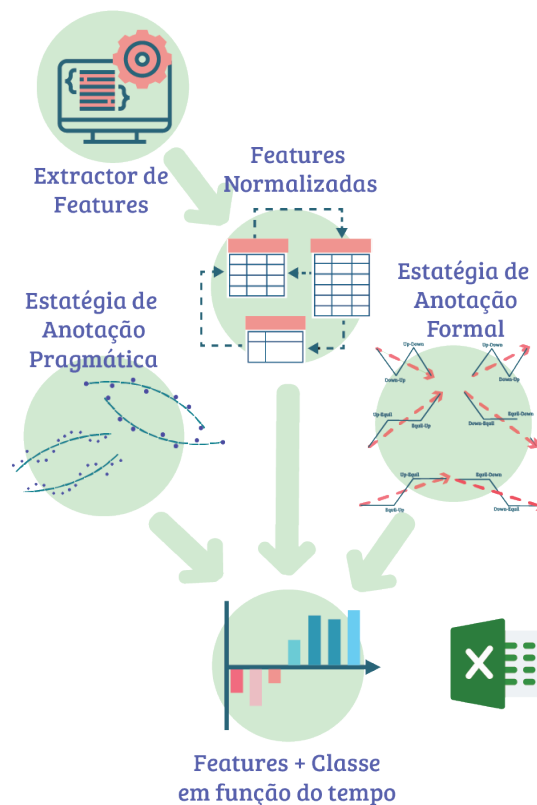
A coleta de dados é realizada para ser utilizada na mineração de dados, pois, é necessária a criação de duas bases de dados, uma para treinamento, responsável por construir o classificador e outra para testes que irá realizar a validação do mesmo.

Os processos de treinamento e classificação preditiva no weka exige que os dados sejam organizados em forma de atributos e cada atributo assemelha-se a uma coluna de uma tabela de um banco de dados, por isso, faz-se necessário uma coleta de dados que seja condizente com esta configuração. Dito isto, o procedimento de extração de dados foi construído da seguinte forma, a primeira etapa foi a coleta das informações ao acessar a plataforma de negociação. Ao realizar este acesso, o softbot ficou encarregado de extrair as informações das cotações do mercado em forma de barras de acordo com o período desejado com a finalidade de construção de base de dados para treinamento e teste. Assim foi possível a construção da base de dados com informações em relação a quantidade de barras, ocorrência dos estados e tempo entre as transições. O arquivo gerado foi em formato txt, mas facilmente convertível para Excel e conseqüentemente lido por quaisquer banco de dados.

## 5.6 Base de dados

Após a caracterização do arquivo dos estados em função do tempo, é preciso realizar a extração das *features* (Comportamentos/Características) das séries e armazená-las em uma base de dados. No presente trabalho as *features* foram convertidas para o formato *.csv* (*comma-separated values*), dada a praticidade e flexibilidade para manipulação dos dados. Este processo envolve uma etapa a parte: a normalização dos dados, detalhada na seção 5.6.1.

Figura 14: Ilustração da etapa base de dados - Excel.



Fonte: Elaborado pelo autor

Com a base de dados normalizada em mãos é crucial relacionar as *Features* obtidas com os elementos que queremos prever, no caso da mineração de dados esse elemento é denominado como classe, ou seja, a classe determina o comportamento do classificador para o mercado. No presente trabalho, as classes definidas para o mercado através das estratégias de anotação, que nada mais é que um conjunto de regras definidas pelo agente supervisor que tenha coerência ao objeto de estudo.

### 5.6.1 Normalização dos dados

Com a base de dados construída, faz-se necessário um tratamento para que ela possa ser interpretada pelo formato arff. Assim a base de dados está disposta em um

formato tabular .csv e com base em uma estratégia de anotação é feita a normalização dos dados.

A normalização dos dados foi feita em forma de análise técnica. Assim a primeira análise técnica começou a implementação de um critério tempo, este critério representa o devido número de ocorrência de barras. Entretanto, o tempo real transcorrido independe do *timeframe* escolhido no terminal de negociação, que pode ser M1, M5, dentre outros, porém, no modelo em curso, o tempo no modelo é independente do *timeframe*, ou seja, quando o *timeframe* está em M1 a contagem é feita em M1 e quando está em M5 é feita em M5 e assim segue os outros *timeframes*.

A análise do tempo também serve para sabermos o tempo entre a transição de um estado para outro e o tempo da transição anterior à atual, como nos é evidenciado na análise markoviana de segunda ordem.

Feita essa alteração, então buscou-se aqui trabalhar com os seguintes atributos categorizáveis, os estados e as transições (última e penúltima) e os atributos numerais. Os tempos entre a última e penúltima transições que são medidos em números de barras para serem independentes do *timeframe*. Por isto, os atributos categorizados pertencem a dois tipos distintos de variáveis, a variável Estado e a variável Transição entre os estados.

Para um melhor manuseio e redução do nível de complexidade do algoritmo dos estados termodinâmicos optou-se por categorizar os estados, com base em suas características em comum. Assim, o número de estados foram definidos em três. Estes três estados foram pensados para abordar a análise de tendências de séries temporais discutidas no capítulo 3.1.

Por conta disto foi estabelecido os seguintes estados para a análise técnica:

1. UPTREND - Tendência de subida.
2. DOWNTREND - Tendência de descida
3. EQUILIBRIUM - Tendência de equilíbrio.

E de acordo com a variável transição entre os estados, tivemos nove categorias possíveis. A representação destas categorias é feita através de duas letras: a primeira letra do estado inicial seguida da primeira letra do estado final na transição, conforme a imagem 15:

Como podem observar na imagem 15, por se tratar de uma transição, quando um estado permanece nele mesmo a transição não existe, ou seja, é representado pelo *null*, nos arquivos de treinamento estas transições não podem aparecer.

Figura 15: Ilustração dos estados e transições.

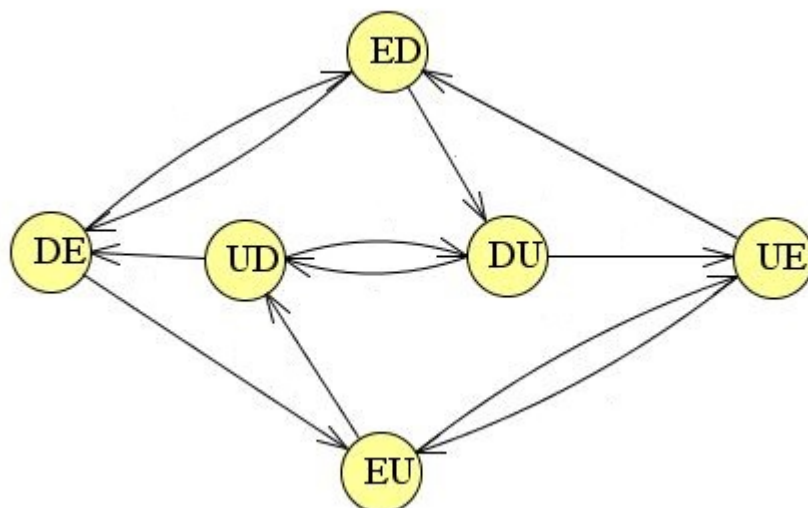
		Estado final		
		U	E	D
Estado inicial	U	NULL	UE	<u>UD</u>
	E	<u>EU</u>	NULL	<u>ED</u>
	D	<u>DU</u>	DE	NULL

Fonte: Elaborado pelo autor.

1. As transições UD e DU significa uma reversão de tendência, ou seja, quando uma tendência de subida muda para uma de descida e vice versa, isso nos indica a realizações de operações de compra ou venda.
2. As transições EU e ED significa uma saída da condição de equilíbrio para a formação de uma tendência, ou seja, nos indica a realização de voltar a realizar operações de compra ou venda.
3. As transições UE e DE significa uma ida a tendencia de equilíbrio, ou seja, a principio não realizar operações, por si só elas não detém de valores preditivos, apenas indica que uma tendência parou de se manifestar.

Também é possível trabalhar com as transições como uma grafo orientado, visualizável na imagem 16.

Figura 16: Ilustração da grafo orientado.



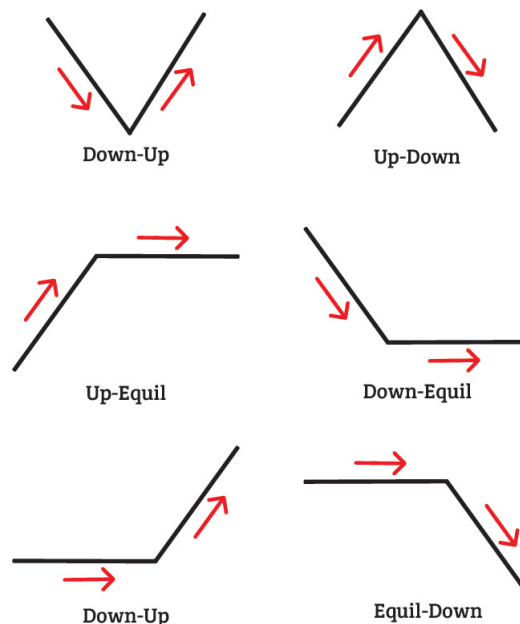
Fonte: Elaborado pelo autor.

A imagem 16 como dito anteriormente representa as transições como uma máquina de estados ou um grafo orientado, onde cada nó representa as transições e as arestas representam os possíveis caminhos que cada transição pode seguir. Graças a esta representação de grafo orientado, é possível deduzir que somente com a análise das transições as séries temporais do Forex atuam de forma cíclica. Então a análise de estados e transições está coadunante com a teoria de séries temporais em relação à sazonalidade e ciclo abordadas no capítulo 3.1.

Como precisamos inserir o elemento tempo para a realização adequada da classificação, pois, sem este elemento é possível que os algoritmos de aprendizado de máquina contidos na mineração de dados aprenda erroneamente como classificar as oportunidades de compra e venda, resultando em *overfitting*. Isto significa que o algoritmo irá considerar apenas as transições e os estados, resultando em um classificador eficiente no sentido de acurácia, pois, ele terá uma alta probabilidade de acerto, mas ineficiente no sentido de uso deste classificador à posteriori, pois ele aprendeu regras simples para classificar por causa do *overfitting*, e isto não combina com as características do mercado que também considera o fator tempo.

Então é necessário analisar as transições em forma de markov de primeira ordem conforme a imagem 17. Onde a inserção no elemento tempo significará o aumento das arestas de cada vetor.

Figura 17: Markov primeira Ordem.

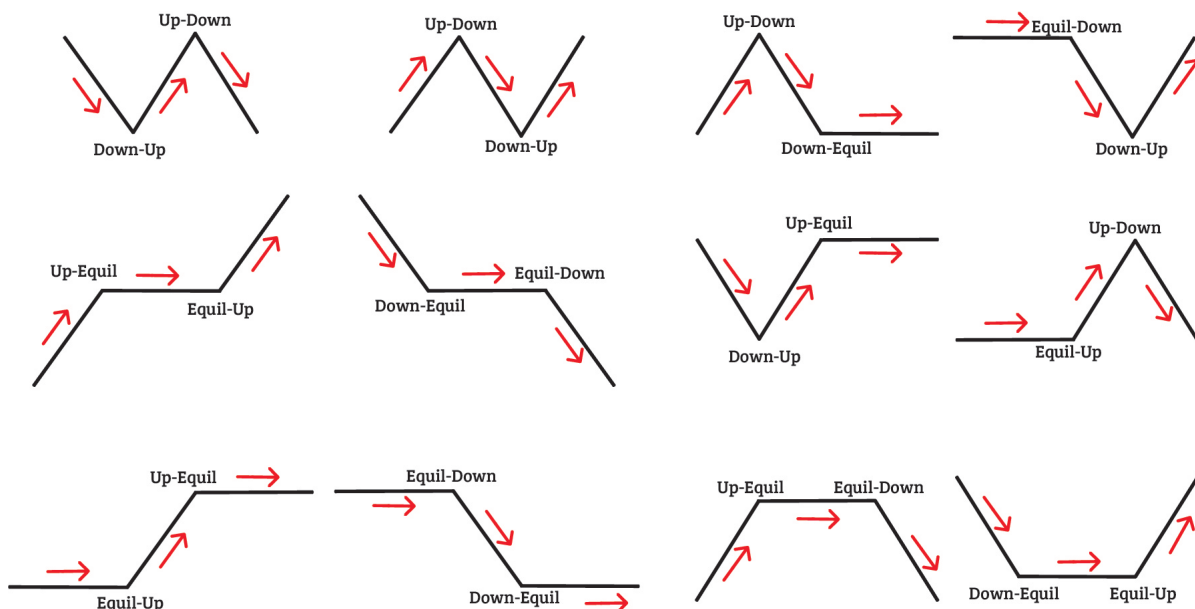


Fonte: Elaborado pelo autor.

Por fim, em consonância um dos objetivos propostos que é a transformação da análise markoviana de primeira ordem e uma análise markoviana de segunda ordem a análise técnica em forma de análise de markoviana de segunda ordem dar-se da seguinte

forma conforme a imagem 18. Assim é possível não só analisar as arestas que representa o tempo, sendo aqui considerado pelo eixo x, mas também a altura, representada pelo eixo y. A altura do padrão é feita através do cálculo da amplitude máxima, este calculo da amplitude máxima.

Figura 18: Markov segunda ordem.



Fonte: Elaborado pelo autor.

Como a presente classificação será feita de forma supervisionada, já que os dados tiveram o suporte de um agente supervisor, ou seja, uma pessoa foi responsável de analisar tecnicamente os dados e classifica-los de acordo com suas características manualmente. Então a base de dados normalizada está disposta da seguinte forma como ilustrado na imagem 19

Figura 19: Ilustração de um recorte dos dados.

UPTREND	DU	UD	6	1 Em Compra
UPTREND	DU	UD	7	1 Em Compra
UPTREND	DU	UD	8	1 Em Compra
UPTREND	DU	UD	9	1 Em Compra

Fonte: Elaborado pelo autor.

Como é possível observar na imagem 19 cada linha da base de dados é considerado uma instancia e as colunas são consideradas atributos desta instância, com exceção da última coluna que representa a classe e é justamente na classe que os algoritmos de aprendizado de máquina ira prever com base nos demais atributos. Que no nosso caso irão prever as operações de compra ou venda a depender dos demais atributos, os estados, as transições e seus respectivos tempos.

Para a definição da classe, foi preciso estabelecer estratégias para defini-las, ou seja, foram elaboradas duas abordagens, na qual estas abordagens reconhecidas como

estratégias de anotação, pois, como dito anteriormente, a metodologia de trabalho empregada foi o aprendizado supervisionado. Sendo assim, a estratégia de anotação é um conjunto de regras pré definido pelo agente supervisor que gera a coluna de classe para cada instância na classificação. O processo aplicado no trabalho é detalhado ao longo do capítulo 6.

### 5.6.2 Weka

Por fim, o último estágio configura-se como a etapa de classificação de fato. Esta etapa utiliza as *Features* combinadas com a classe, ou seja, todas as instâncias estão rotuladas. Essas instancias serão utilizadas para a construção do classificador e criação de arquivos "comparadores", ou seja, arquivos que já tenha as respostas das previsões. Estas comparações geram métricas e resultados necessários para as avaliações dos resultados.

Figura 20: Ilustração da etapa do Weka.



Fonte: Elaborado pelo autor.

## 6 Definição e construção do classificador.

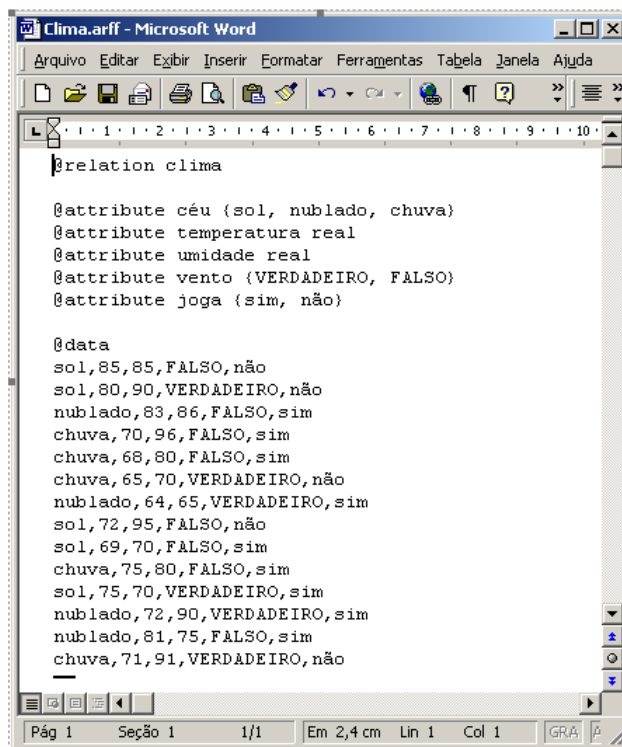
Como abordado nos capítulos anteriores, o grande objetivo deste projeto é a realização da classificação correta das operações do mercado, logrando uma confiança com uma taxa de probabilidade aceitável para a operação deste modelo em uma plataforma real de investimento no mercado Forex.

A principal ferramenta para a realização da classificação é o WEKA. O Weka (*Waikato Environment for Knowledge Analysis*) é uma solução de código aberto implementado na linguagem de programação Java, que tem como principal objetivo trazer um conjunto de algoritmos e técnicas de mineração de dados (SOUTO, 2004). A motivação da escolha desta ferramenta é justificada, pois, ela contém vários algoritmos de mineração de dados e permite que os usuários comparem um ou mais algoritmos. Além de oferecer uma interface intuitiva, ser gratuito e ter disponíveis muitas referências em relação ao seu uso (MANHÃES et al., 2011).

Os dados que serão interpretados para a realização da classificação devem estar contidos em um arquivo Arff (*Attribute-Relation File Format*). Este formato é um formato proprietário do Weka, é nele que devemos descrever o domínio do atributo para o Weka interpretar automaticamente seus valores (SOUTO, 2004). Sendo assim, o conteúdo do arquivo Arff deve estar disposto da seguinte maneira: A primeira parte consiste a lista de todos os atributos, ou seja, devemos estabelecer o seu tipo aqui. A segunda parte deve conter todos os valores que serão instanciados para a classificação (SOUTO, 2004). A título de ilustração segue a imagem 21 exemplificando um arquivo Arff, pode-se notar que os atributos são similares a tipo de variáveis e os dados podem ser considerados como tupla de banco de dados.

Configurado o arquivo Arff corretamente, será utilizado o modelo de classificação propriamente dito do Weka. Segundo Gonçalves (2007) este modelo é caracterizado pela existência de duas fases: O aprendizado e o teste. Na primeira fase, a de aprendizado, é aplicado um algoritmo classificado sobre um conjunto de dados de treinamento e nos é retornado a construção do classificador. Após a construção deste classificador, é iniciado a fase de testes, é nesta fase que é avaliado a acurácia do modelo através do conjunto de dados de teste. Esta acurácia indica a porcentagem de observações feitas corretamente pelo modelo classificador, ou seja, nos é mostrado aqui a eficiência do modelo, se ele serve para classificar novos casos ou não.

Figura 21: Exemplo de um arquivo Arff genérico.



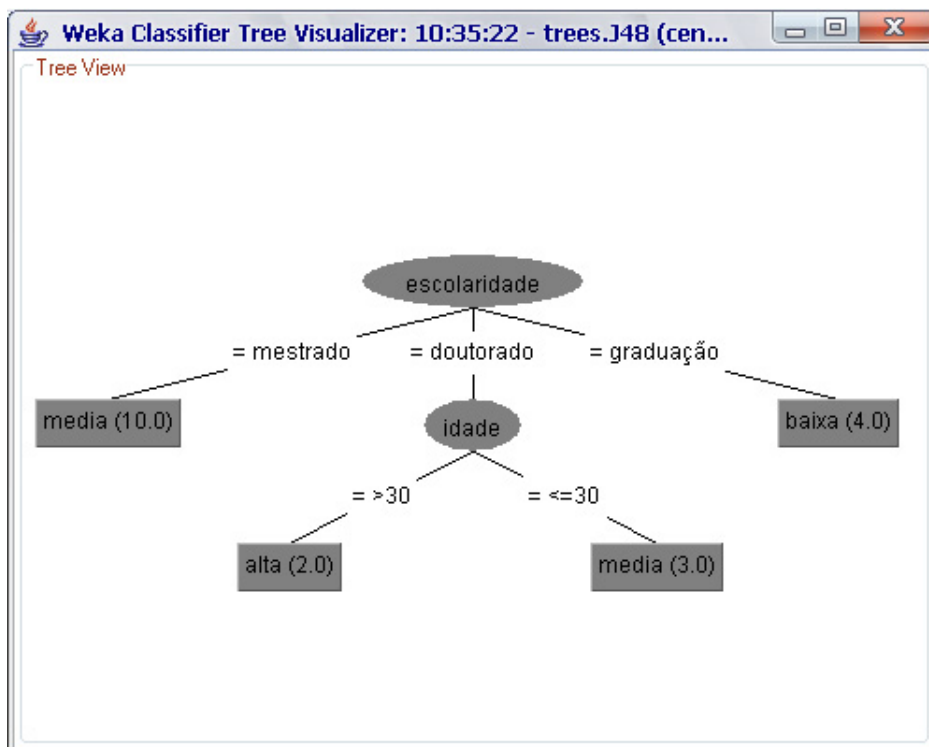
Fonte: Elaborado pelo autor.

## 6.1 Classificadores

Os algoritmos de classificação utilizados neste trabalho foi o classificador baseado em regras ZeroR e a árvore de decisão J48. Ambas as técnicas são de aprendizagem de máquina supervisionada, elas encontram-se dentro da mineração de dados abordado no capítulo 4.3. Além disso, foram escolhidas por conta de serem capazes de auxiliarem na construção de sistemas complexos por meio de conhecimento de forma automática, ou seja, métodos que se baseiam em experiências acumuladas para a realização de predições futuras (RODRIGUES, 2007).

Não obstante, as técnicas de árvore de decisão apresentam-se de forma eficiente na construção de classificadores preditivos, por conta da sua estratégia de divisão e conquista. Por isso, um problema complexo pode ser dividido em sub problemas relativamente simples, assim é possível reduzir a complexidade e facilitar a resolução do problema em questão. Sendo assim, esta técnica de árvore trata de uma análise estatística que se apropria da divisão dos conjuntos de exemplos fornecidos em diversos subconjuntos, que são utilizados em treinamento. Com a finalidade de que cada subconjunto pertença a uma classe de forma singular, ou que esta seja predominante, até que não seja necessárias novas divisões.

Figura 22: Exemplo de uma árvore plotada com o algoritmo j48.



Fonte: Elaborado pelo autor.

Como podem observar na imagem 22, na prática a árvore gerada é composta de nós, similar a uma árvore binária amplamente estudada na computação. Estes nós é a representação dos atributos. Outro elemento que compõe a árvore são os ramos, nos quais eles são responsáveis por receber os possíveis valores para os atributos selecionados. Também é factível visualizar o seu caminho do início ao fim através das folhas e é este caminho o responsável por gerar a regra classificatória.

### 6.1.1 Especificação da técnica

Como explanado anteriormente, a técnica utilizada para a classificação foi a árvore de decisão e o classificador ZeroR.

O uso do classificador ZeroR se deu por questões que ele serve como uma linha de base, pois, ele realiza as predições de acordo com a classe majoritária dos exemplos de treinamentos. Ele é considerado um classificador padrão, apesar de realizar esta análise trivial considerando apenas a moda dos atributos, ele serve para comparar o desempenho dos algoritmos mais complexos, ou seja, se um algoritmo mais complexo tiver uma acurácia mais baixa do que o classificador ZeroR dado um conjunto quaisquer de dados, isto significa que o algoritmo escolhido não é adequado para resolver o problema proposto, ou que o conjunto não contém informações uteis para a extração de conhecimento (SOUZA, 2010).

Para a realização da classificação preditiva de fato foi escolhido o classificador J48, pois, é um algoritmo que determina a criação da árvore de forma top-down, ou seja, o atributo mais significativo é generalizado em relação aos outros atributos do conjunto, definindo a raiz da árvore. Após a definição da raiz da árvore os próximos nós serão o segundo atributo mais significativo e assim sucessivamente, até a montagem total da árvore (RODRIGUES, 2007).

### 6.1.2 Efeito Comparativo

Para realizar um estudo comparativo entre as técnicas de classificação foram adicionados mais três algoritmos, estes algoritmos trazem técnicas de mineração de dados diferentes das técnicas citadas anteriormente. Sendo assim, foram adicionados no presente trabalho o classificador baseado em redes neurais *Multilayer Perceptron*, o classificador bayesiano *Naive Bayes* e o classificador baseado no vizinho mais próximo *KStar*.

É importante ressaltar que a escolha destes novos classificadores justifica-se, pois, a mineração de dados aborda uma variedade de técnicas, então não é interessante manter o trabalho limitado apenas à dois classificadores, dado a complexidade do mercado Forex, então é interesse a busca de outras alternativas que busquem fundamentar, ou até refutar o uso da técnica principal que é abordada na seção 6.1. Como estes algoritmos foram utilizados para efeito comparativo, os testes realizados com estes classificadores supracitados tiveram a sua configuração *default* no WEKA.

O classificador *KStar* é fundamentado na abordagem de vizinhança mais próxima, ou seja, ele busca encontrar todos os exemplos que treinamento que seja semelhante aos atributos do exemplo de teste. Sendo assim, este classificador baseado em exemplos utiliza-se das instâncias que tenha características semelhantes através de uma função de similaridade (SOUZA, 2011).

O classificador *Multilayer Perceptron* é baseado no algoritmo *Backpropagation* para realizar a classificação das instâncias, ou seja, é uma rede neural artificial que se utiliza da retro-propagação para processar seus elementos. Sendo assim, é criada uma rede neural que consiste em uma função de ativação não linear na camada oculta, que busca mapear os valores de entrada e saída, ou seja, busca padrões adaptativos de formas não lineares de acordo com a base de dados. A sua estrutura básica é formada por uma arquitetura de três camadas, uma camada de entrada que define as instâncias, uma camada intermediária (oculta) que define as funções matemáticas e uma camada de saída que define as instâncias devidamente classificadas (SAHOO; KUMAR, 2012; ARORA, 2012).

Por fim, os classificadores bayesianos vem como forma de abordagem não determinística fundamentada na inferência probabilista, ou seja, a classe não é prevista com total certeza. Este classificador é responsável por estimar as classes a serem previstas utilizando

o modelo probabilístico de Naive Bayes. Ele realiza a sua inferência através de um estimador de *Kernel* para os atributos dos dados, calculando uma distribuição normal com o uso da média e desvio padrão (SOUZA, 2011).

Sendo assim, o total de algoritmos utilizados no trabalho proposto pode ser ilustrado na tabela 2

Tabela 2: Técnica de classificação e algoritmos de aprendizado de máquina utilizados.

Técnica de Classificação	Algoritmo de Classificador
Árvore de Decisão	J48
Baseado em regras	ZeroR
Bayesiano	NaiveBayes
Vizinho Mais Próximo	KStar
Redes Neurais Artificiais	Multilayer Perceptron

## 6.2 Abordagem pragmática e formal.

Para a realização da classificação dos dados foram utilizadas duas abordagens preditivas baseadas no modelo markoviano de segunda ordem. Ambas as abordagens têm como o principal objetivo a otimização de uma estratégia classificatória, ou seja, a construção de um classificador eficiente que consiga prever corretamente as tendências futuras do Forex.

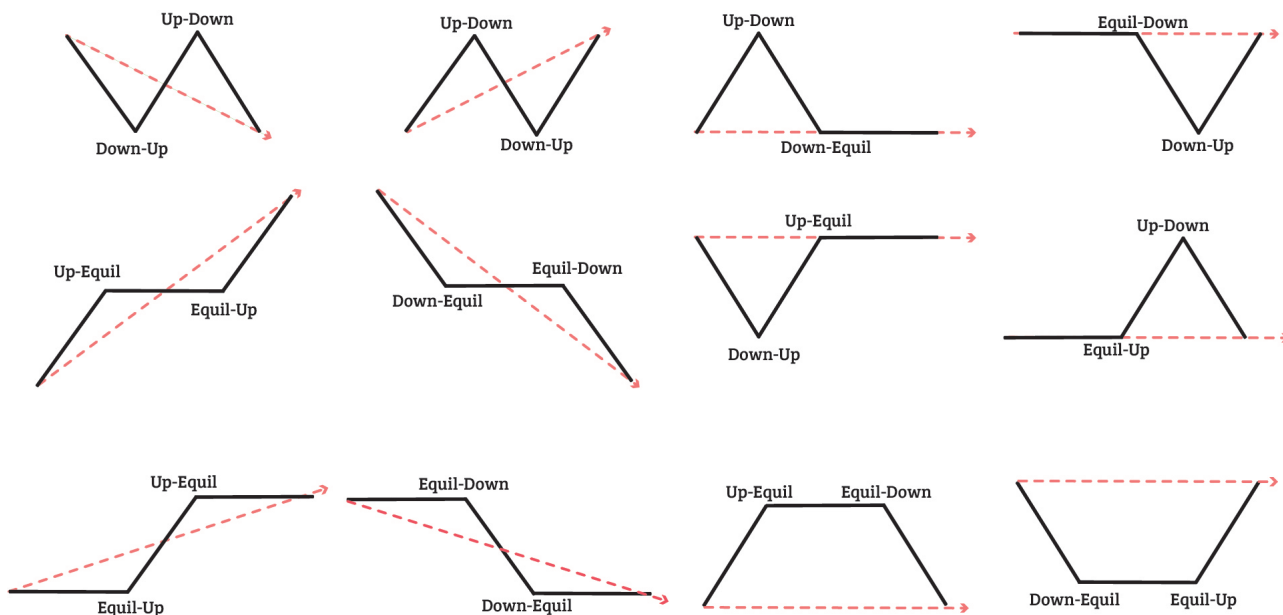
Estas estruturas são capazes de gerar regras por intermédio dos algoritmos de classificação. Sendo assim, a resultante das regras formadas por estas estruturas consideram um conjunto de elementos: estados, transições, tempo de ocorrência das barras, amplitude e classes (elemento preditivo: comprar, vender ou não operar).

### 6.2.1 Formal

Com base no modelo markoviano de segunda ordem foi possível extrair uma análise formal dos estados e suas transições, considerando todas as variáveis para a realização da predição, exceto a amplitude. A criação deste modelo veio a luz por conta da necessidade de superação do sobre-ajuste, pois, ao desenvolver um classificador, mesmo quando seus dados estão normalizados, o classificador pode acabar inferindo regras simples, dando ao agente supervisor uma falsa impressão de predição.

Como podem observar, com base na figura 23 são apresentados todos os conjuntos de estados e transições representado pelos traços em preto e o elemento preditor, ou seja, a classe representada na seta tracejada em vermelho. Sendo assim é possível inferir que dado um conjunto de estados e transições o classificador irá prever com base na classe, esta classe representada por uma tendência, seta de subida, seta de descida e seta estagnada.

Figura 23: Representação da Abordagem Formal.



Fonte: Elaborado pelo autor.

As setas de subida e descida são realizadas operações de compra ou venda, compra para as de subida e venda para as de descida. Em contraponto, as setas estagnadas não serão realizadas operações.

Não obstante, a tabela 3 amostra quais são os estados-transições demonstrados na figura 23 que deverão ser levando em consideração para as realizações de operações no Forex, os demais estados não serão considerados para a realização de operações. Esta abordagem formal foi desenvolvida para possibilitar a execução de uma análise markoviana de segunda ordem em forma de uma família de estados, na qual esta família considera os estados-transições e o tempo entre as transições. Assim foi possível a construção de uma regra que englobasse todas as características do modelo markoviano, ou seja, todos os dados propostos pelo modelo são considerados para a realização da previsão, porém, de forma teórica.

Tabela 3: Estados-Transições validos para a realização de operações.

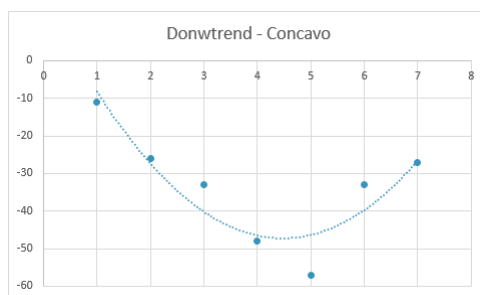
Estado-Transição	Classe
Down-up, Up-Down	Vender
Up-Down, Down-Up	Comprar
Up-Equal, Equil-Up	Comprar
Down-Equal, Equil-Down	Vender
Equil-Up, Up-Equal	Comprar
Equil-Down, Down-Equal	Vender

## 6.2.2 Pragmática

Para a definição de uma estrutura que se aproxima do modelo real de fato, envolvendo lucros e perdas foi necessário a criação de uma abordagem pragmática. Esta abordagem ficou encarregada de analisar as tendências de subida (*uptrend*), descida (*downtrend*), que ultrapassa um certo limiar dado em pips pré definido. A tendência de estagnação (*equilibrium*) não foi analisada, pois, é uma tendência na qual o mercado está inerte, então os possíveis lucros nestas tendências não são relevantes quanto nas outras, e por se tratar de um elemento indefinido, pode ocasionar falsos positivos, sendo assim não é realizado operações bem como não são realizadas análises. As figuras 24, 25, 27, 26 representa graficamente a análise pragmática. Esta estrutura é pautada na consideração da amplitude máxima (para as tendências de subida) e mínima (para as tendências de descida). No que lhe concerne, esta amplitude é definida com base na quantidade de pips extraído da série temporal, no caso a amplitude é o valor máximo em pips de uma determinada tendência.

Hipoteticamente uma tendência tem os seguintes valores ordenados: 1,4,5,6,7,4,2,3. Como os valores desta tendência estão positivos esta tendência hipotética será considerada de subida a título de ilustração. Sendo assim a sua amplitude máxima é o valor 7, o valor inicial 1 e o valor final 3. A direção da tendência que transpõe os dados são estabelecidos por intermédio da regressão linear quadrática, assim é possível observar linhas de tendências concavas e convexas nesta estrutura. A importância desta estrutura justifica-se na questão de eliminar tendências que demonstre valores de prejuízo ou lucro insuficiente por conta do limiar, nestes casos não são realizados operações. Também com esta abordagem é possível identificar falsos positivos nas tendências estes falsos positivos é onde deve ser dado uma atenção maior, já que são realizadas operações, porém, uma operação que deveria comprar dando um lucro  $x$  pode ocasionar em um prejuízo  $y$ .

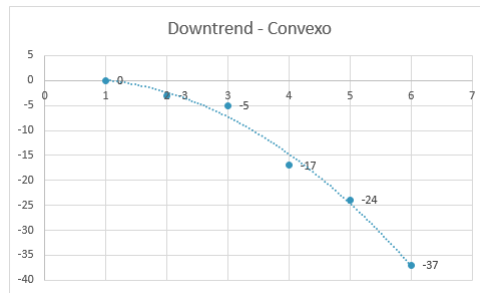
Figura 24: Gráfico de tendência Downtrend Concavo



Fonte: Elaborado pelo autor.

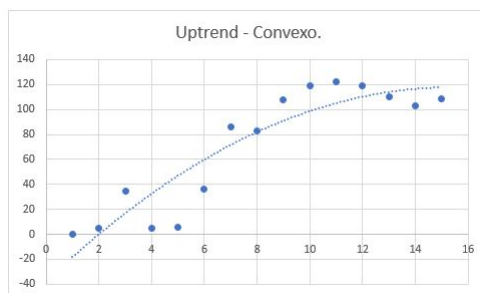
- A figura 24 ilustra uma tendência de descida de forma concava na qual o valor da amplitude mínima é menor que valor final da tendência e o valor final da tendência é menor que o valor inicial da mesma.

Figura 25: Gráfico de tendência Downtrend Convexo.



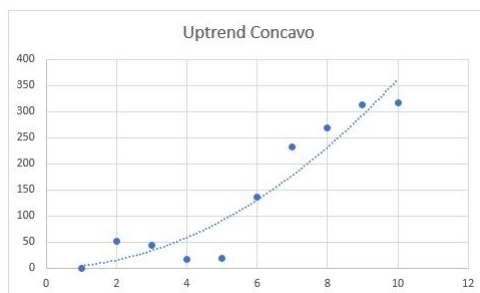
Fonte: Elaborado pelo autor.

Figura 26: Gráfico de tendência Uptrend Convexo.



Fonte: Elaborado pelo autor.

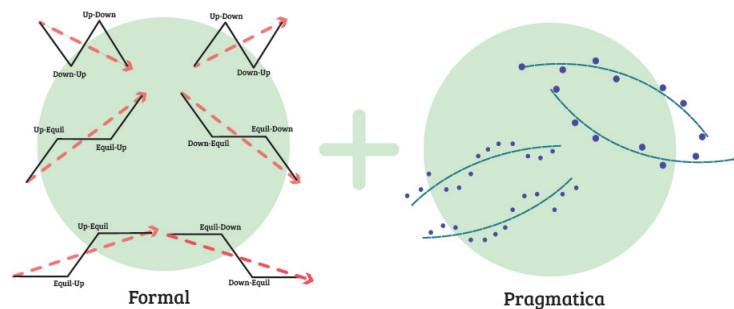
Figura 27: Gráfico de tendência Uptrend Concavo.



Fonte: Elaborado pelo autor.

- A figura 25 representa uma tendência de descida de forma convexa onde o valor da amplitude mínima corresponde ao valor final da tendência e por sua vez o valor final da tendência é menor que o valor inicial.
- A figura 26 representa uma tendência de subida de forma concava na qual o valor da amplitude máxima é maior que o valor final e inicial da tendência (e o ultimo valor da tendência é maior que o seu valor inicial)
- E por fim a figura 27 ilustra uma tendência de subida de forma convexa, onde o valor da amplitude máxima é igual ao valor final da tendência.

Figura 28: Conceito da abordagem híbrida.



Fonte: Elaborado pelo autor.

### 6.2.3 Híbrida

A figura 28 ilustra um conceito da fusão entre as duas abordagens citadas nas seções anteriores (6.2.1 e 6.2.2), resultando em uma abordagem híbrida.

A abordagem híbrida é um estudo estatístico da correlação entre às duas anotações supracitadas anteriormente, ou seja, ela é pautada na realização da análise dos padrões de segunda ordem que tiveram "match" com as tendências da anotação pragmática. A realização desse estudo justifica-se, pois, através desta abordagem é possível comprovar a utilidade prática da abordagem formal que é baseada no modelo dos três estados.

Para a criação desta análise primeiramente foi observado empiricamente as colunas referente à classe de ambas as abordagens. Posteriormente é verificado quais são as classes das famílias de estados de segunda ordem que se correlacionam com as classes tendências em pips. Sendo assim, foi estabelecido duas regras:

- Caso as correlações entre as anotações sejam iguais a classe de ambas as abordagens permanecem, ou seja, se a classe da abordagem formal definir uma operação de compra e a classe da abordagem pragmática também definir uma operação de compra, então a operação de compra prevalece (a mesma coisa acontece para as operações de venda e não operação).
- Caso as correlações entre as anotações sejam diferentes a classe será definida para não realizar operações, ou seja, se a classe da abordagem formal definir uma operação de compra e a classe da abordagem pragmática definir não operar (ou vice-versa, poderia uma abordagem definir compra e a outra venda), então não será realizado operações.

## 7 Experimentos, Testes e Resultados

Os experimentos adotados no trabalho teve como o principal objetivo obter conhecimentos extraídos da classificação da base de dados normalizadas no modelo markoviano de segunda ordem, utilizando tanto a abordagem formal quanto a abordagem pragmática. A condução do presente ensaio já fora estabelecido através da metodologia abordada na seção 5.2. Para a aplicação dos algoritmos de aprendizado de máquina foi utilizado a ferramenta weka e para efeito de comparação com o classificador em árvore J48 foram realizados experimentos com os algoritmos BayesNet, KStar, ZeroR e MultilayerPerceptron para realizar um confronto de desempenho.

### 7.1 Harmonização dos datasets

O primeiro estudo realização diz respeito a harmonização<sup>1</sup> da base de dados proposta, em especial este procedimento foi utilizado para definir a escolha do limiar. A definição do limiar é de suma importância, pois, ele define a quantidade de exemplos para treinamento e testes na abordagem pragmática. Para tal, é necessário estabelecer uma quantidade de exemplos suficientes para que os classificadores possam generalizar de forma suficiente os dados.

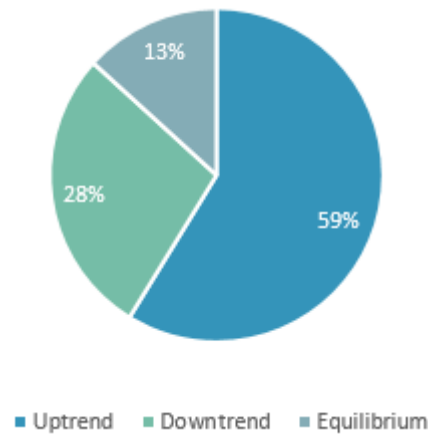
Os dados deste estudo foi sistematizados em forma de gráfico de pizza para melhor visualização, foi realizado um estudo preliminar utilizando à série temporal do mês de junho como referência. A série em questão possui 797 instância de *equilibrium*, 1680 instâncias de *downtrend* e 3530 instâncias de *uptrend*, é possível observar suas porcentagens na figura 29.

Posteriormente foi experimentado empiricamente alguns limiares para a definição das classes. Foi obtido o seguinte resultado como ilustrado na figura 30. Foi possível observar que à medida que aumentamos o limiar para definir a classe na abordagem pragmática, a quantidade de exemplos diminui para a realização de operações de compra ou venda. Isso se justifica, pois, a medida que aumentamos o limiar, existem poucas barras (no caso desse mês em específico) que tenha valores maiores que 25 pips, e no momento que vamos aumentando, existe menos exemplos ainda. Até chegar em um momento, no caso de um limiar de 200 pips, que não exista quase nenhuma oportunidade de venda e muito poucos casos para venda, apenas 4%

Sendo assim foram realizados uma bateria de testes para encontrar o mês com a base de dados mais harmonizada. Também buscou-se aqui diminuir o valor do limiar para

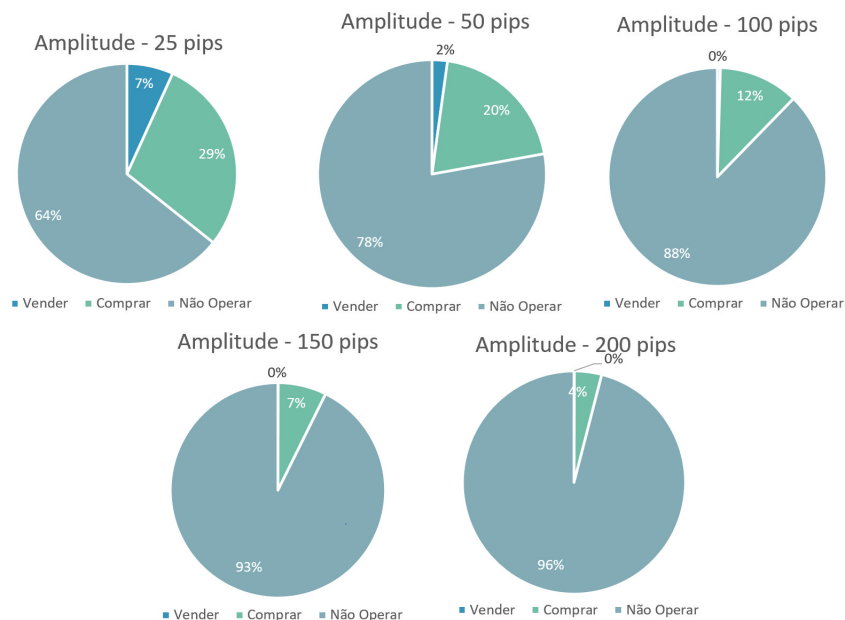
<sup>1</sup> A harmonização diz respeito ao equilíbrio e simetria da base de dados.

Figura 29: Porcentagem da quantidade de instâncias da série temporal - Junho 2018.



Fonte: Elaborado pelo autor.

Figura 30: Porcentagem de classes de acordo com os limiares pré definidos - Junho 2018.

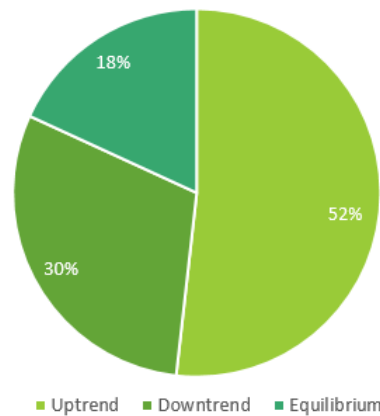


Fonte: Elaborado pelo autor.

poder obter mais exemplos de classe para as operações de compra e venda. Após a bateria de testes o mês considerado com o maior grau de harmonização foi o mês de outubro de 2018. Como é possível observar na figura 31. Apesar das porcentagens dele serem parecidas com o de junho de 2018, a diferenciação se da, na quantidade de instâncias, pois, neste mês a quantidade de instâncias para *equilibrium* e *downtrend* são maiores que a do mês de junho e há uma diminuição nas instâncias de *uptrend*.

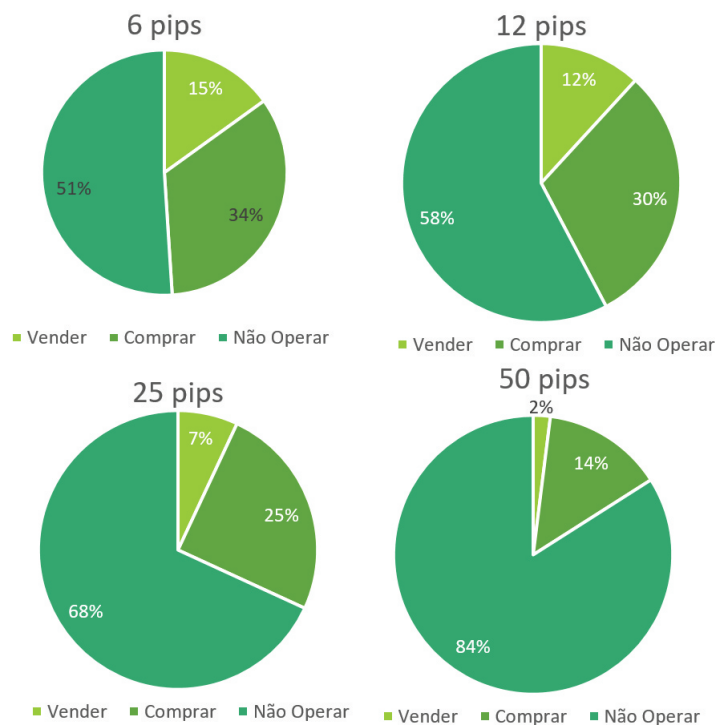
Bem como na figura 32 é possível observar que a diminuição do valor dos limiares também contribuíram para uma maior generalização dos exemplos, equilibrando, não em sua totalidade, as operações de compra e venda.

Figura 31: Porcentagem da quantidade de instâncias da série temporal - Outubro 2018.



Fonte: Elaborado pelo autor.

Figura 32: Porcentagem de classes de acordo com os limiares pré definidos - Outubro 2018.



Fonte: Elaborado pelo autor.

## 7.2 Treinamento e Teste

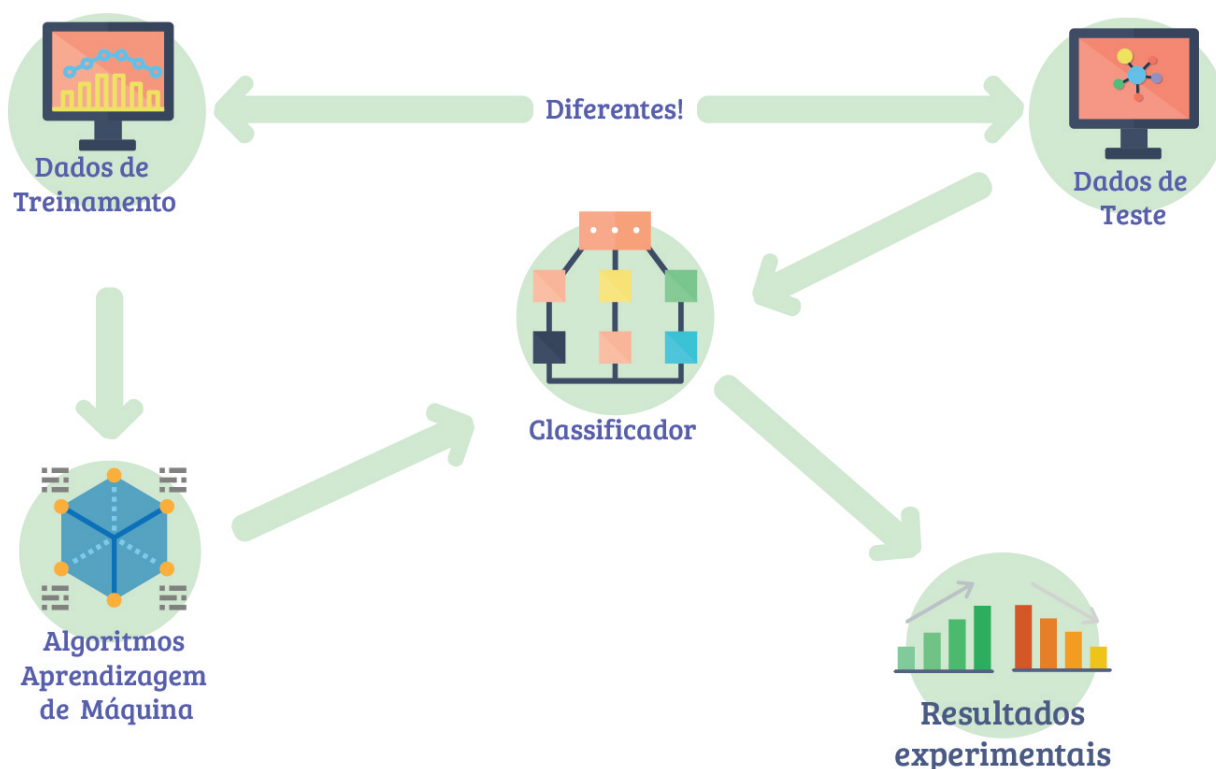
O treinamento de classificadores se dá através do processamento da base de treinamento e uma base diferente para a realização dos testes. De acordo com [Bortoli \(2015\)](#) os conjuntos de treinamento não podem ser grande demais para não gerar superadaptação dos dados e nem miseráveis demais para não gerar um desempenho ruim no classificador. Por isto é necessário tomar cuidado, pois, o desempenho do conjunto de treinamento pode enganar o agente supervisor, caso seja utilizado os mesmos dados para treinamento e teste, porque assim o algoritmo memoriza a base de dados e não gera conhecimento

novo. Por isto neste trabalho é usado uma base de dados para generalizar as abordagens obtendo assim um classificador amplamente aplicável. Os testes foram realizados com um conjunto de dados independentes.

Este processo de classificação foi ilustrado na figura 33 para um melhor entendimento. Logo é possível concluir que, os dados de treinamento devem ser diferentes dos dados de testes. Os dados de treinamentos serão processados através dos algoritmos de aprendizagem de máquina, gerando um classificador. Posteriormente os dados de testes serão processados pelo classificador anteriormente produzido, gerando assim os resultados experimentais.

Coadunante com o que foi dito anteriormente no presente trabalho foram utilizadas bases de ambas as abordagens supracitadas na seção 6.2. As bases de dados tiveram como referência as séries temporais extraídas dos *ticks* modelados do Forex com referência do mês de outubro de 2018 para treinamento, com o objetivo de prever os meses de 2018, mesmo prevendo meses antecessores, o que importa de fato são os dados das séries temporais, já que no teste o algoritmo não saberá qual é a classe, assim a ordem da inversão dos meses não influencia de forma negativa para a realização das previsões.

Figura 33: Ilustração do funcionamento do classificador.



Fonte: Elaborado pelo autor.

A utilização de uma base de dados com uma referência mensal justifica-se, pois, é possível extrair uma média de 6000 instancias de dados que por sua vez estas instâncias contemplam toda a família de estados adotadas na abordagem formal, ou seja, é possível

identificar todos os padrões ilustrados na figura 18. No caso da abordagem pragmática, foi necessária a criação de bases de dados distintas, cada uma representando um limiar 6pips até 50 pips.

## 7.3 Análise dos Resultados

Após a realização dos experimentos com treinamento e testes das abordagens formal e pragmática, buscou-se executar o classificador com o propósito de obter as acurácias dos classificadores de modo a realizar uma comparação entre as abordagens e também uma comparação entre a eficiência entre os diferentes classificadores. Buscando obter o classificador mais adequado para realizar as operações no mercado, este fator foi determinado por meio do estudo da qualidade das previsões. Pois, não basta apenas o classificador demonstrar a melhor acurácia, mas sim, apresentar ser um classificador viável em termos de implementação, manutenção, custo computacional e capacidade de generalização. O presente trabalho focou-se na acurácia e capacidade de generalização, apesar de não envolver todos os atributos supracitados, com estes atributos foi possível realizar inferências em relação à viabilidade e eficiência dos classificadores.

### 7.3.1 Análise do experimento formal

Como possível observar na tabela 4 é ilustrado as porcentagens de acurácia do classificador, esta acurácia é obtida com total de instância em relação à quantidade de instância corretamente classificadas, ou seja, são estas instancias que realizam predições assertivas. Não obstante, a tabela 5 representa as matrizes de confusão. É importante analisar a matriz de confusão, pois, ela demonstra de fato o que aconteceu com a base de dados.

Segundo Carrijo (2004) uma matriz de confusão é a consideração de uma hipótese quaisquer, demonstrando a realidade de todas as instâncias classificadas, ou seja, o número de classificações corretas confrontando o número de classificadores que de fato foram preditas para cada classe através de um conjunto de dados. Logo é possível inferir a análise da matriz de confusão do presente trabalho da seguinte forma: coluna a,b,c representando as classes e os elementos que são as instâncias dispostas em forma de matriz.

Tabela 4: Resultado experimento formal

Classificador	Instâncias corretas	Instâncias incorretas	acc%
J48	5643	681	89,2
NaiveBayes	4514	1810	71,4
KStar	5567	757	88,1
MultilayerPerceptron	5587	737	88,3
ZeroR	4047	2277	63,9

Como dito anteriormente, a tabela 4 apresenta os classificadores utilizados para a realização da comparação, é possível observar que o classificador J48 teve o melhor desempenho entre eles. Outra questão interessante é todos os classificadores tiveram um desempenho melhor que o classificador base ZeroR, então a princípio é possível escolher quaisquer um desses classificadores para realizar operações com base nessa abordagem no Forex. Bem como também é possível observar que o NaiveBayes obteve a segunda colocação no quesito acurácia da modelagem de dados, e apenas os outros três classificadores concorrem entre si por conta dos seus valores sempre bem próximos. Entretanto, ao analisar suas matrizes de confusão é possível observar alguns desbalanceamentos nos dados que serão abordados a seguir.

Tabela 5: Matriz de confusão

Técnica de Classificação	Algoritmo	Comprar	Vender	Não Operar
Árvore de Decisão	J48	a	b	c
		3952	61	34
		296	1398	71
		134	84	293
Bayesiana	NaiveBayes	a	b	c
		2977	713	357
		337	1392	37
		92	274	145
Vizinho Mais Próximo	KStar	a	b	c
		4025	22	0
		385	1379	2
		205	143	163
Redes Neurais Artificiais	Multilayer Perceptron	a	b	c
		3908	136	3
		256	1503	7
		159	176	176
Baseado em regras	ZeroR	a	b	c
		4047	0	0
		1766	0	0
		511	0	0

O classificador J48 apresenta uma matriz de confusão com um alto grau de confiança, pois, das instâncias classificadas para a realização de operações de compra (4195) ele errou 69 (1,5%), pois, são as operações que deveriam ser realizadas compras e no caso foram realizadas vendas, neste caso realmente há um prejuízo por conta da realização dessa operação inversa. Os casos de não operação são irrelevantes, pois, o resultado é neutro, não realizar operação quando se deveria comprar não resulta em prejuízo ou lucro.

Por outro lado, já quando é necessário vender é possível perceber um aumento no caso da operação inversa, pois, a um erro 296 (21%) instâncias.

As matrizes de confusão dos classificadores *KStar* e *Multilayer Perceptron* apresentam um comportamento semelhante ao do classificador J48, bem como as suas acurácias. Entretanto, o algoritmo *KStar* foi o que mais se destacou para as operações de compra, pois, ele conseguiu classificar corretamente o maior número de casos, errando apenas 22 casos. Já quando ele realiza as operações de venda o número de erros dele é maior que o J48. A mesma coisa acontece com o classificador *Multilayer Perceptron* ele acaba sendo melhor para as operações de venda, porém, o inverso acontece nas operações de compra.

Então um possível critério de desempate em relação à escolha da implementação do classificador seria a complexidade de implementação desses classificadores no Forex, ficando a critério de desenvolvedor responsável. Já no algoritmo *NaiveBayes* é possível observar que ele é o único na qual a acurácia aproxima-se do desempenho do ZeroR bem como sua matriz de confusão erra consideravelmente as operações de compra, mas tem um desempenho semelhante ao J48 nas operações de venda.

### 7.3.1.1 Concordância.

Fundamentado no resultado da matriz de confusão, foi calculado um fator de concordância conhecido na estatística como coeficiente de Kappa. Este fator de concordância foi utilizado para atestar a generalização do modelo no tocante a construção do classificador perfeito.

O coeficiente de Kappa é uma medida estatística utilizada para medir a concordância entre dois ou mais observadores. Na classificação os observadores são representados pela classe. Sendo assim é possível obter a concordância real das instâncias classificadas. Os valores deste coeficiente devem ser interpretados em uma escala que vai de  $<0$  (Nenhuma Concordância) e 1 Concordância perfeita. Os demais valores são interpretados da seguinte forma: 0 até 0,20 concordância leve, 0,21 até 0,40 concordância razoável, 0,41 até 0,60 concordância moderada, 0,61 até 0,80 concordância forte ou substancial, 0,81 até 1 concordância perfeita ou quase perfeita (SOUZA, 2011).

Tabela 6: Resultado dos coeficientes Kappa e sua interpretação.

Classificador	Kappa	Interpretação
J48	0,7712	Concordância Forte
NaiveBayes	0,4733	Concordância Moderada
KStar	0,7413	Concordância Forte
MultilayerPerceptron	0,7572	Concordância Forte

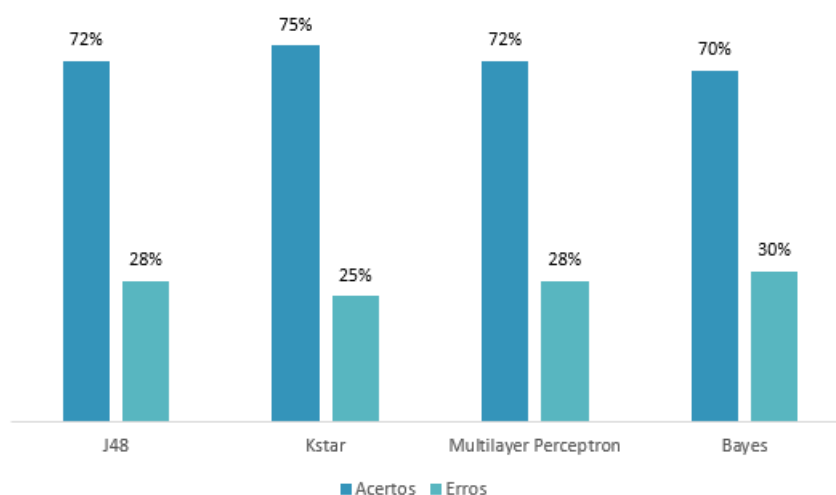
A tabela 6 apresenta os resultados dos coeficientes obtido de cada classificador. Observa-se que o único classificador que não obteve uma concordância satisfatória foi o

*NaiveBayes*, porém, nenhum dos outros três tiveram o grau de concordância máximo. O classificador J48 apresentou o maior coeficiente de concordância dentre os demais.

### 7.3.1.2 Previsões

Com base na análise de resultados dos classificadores explanados anteriormente. Fez-se necessário uma validação no que diz respeito a capacidade preditiva dos classificadores propostos. Dito isto, com os classificadores devidamente treinados foram verificados o percentual de acerto das previsões em função da classe pré definida nos dados de testes. A título de ilustração, foram criados dois *datasets* de testes, um contendo a classe e outro não.

Figura 34: Porcentagens do valores preditos utilizando os *datasets* de teste.



Fonte: Elaborado pelo autor.

Após a realização do treinamento, foi realizado o teste com o *dataset* que não continha a classe (já que é o elemento que queremos prever), o classificador retornou os elementos que foram previstos e a fins de validação acadêmica foi feito uma comparação entre o arquivo gerado pelo classificador e o arquivo que já continha as classes. Os resultados podem ser observados no gráfico ilustrado na figura 34

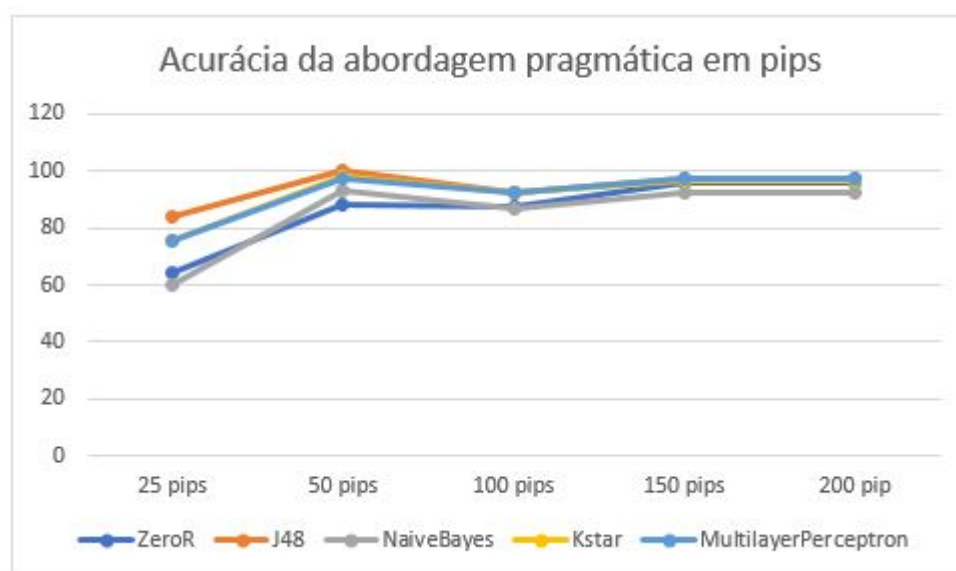
Com esta figura (34) é possível observar que o classificador KStar obteve o melhor desempenho entre os demais classificadores, apesar desde desempenho não ser tão superior aos demais, ele também não é irrisório, dado que qualquer erro de operação do mercado Forex pode impactar em perdas imensuráveis a depender do lote apostado.

### 7.3.2 Análise do experimento pragmático

Bem como no estudo do experimento formal, primeiramente foi feito um estudo preliminar no experimento pragmático. Com base nos resultados experimentais do experimento pragmático a figura 35 (representando uma plotagem da tabela 7) apresenta a

evolução da acurácia da abordagem pragmática em relação ao aumento do limiar em pips. Como é possível observar, quando a quantidade de pips do limiar dos *datasets* aumentam os dados começam a se estabilizar de forma crescente.

Figura 35: Evolução da acurácia da abordagem pragmática.



Fonte: Elaborado pelo autor.

Tabela 7: Resultado das acurácias do experimento pragmático.

Classificador	acc 25 pips	acc 50 pips	acc 100 pips	acc 150 pips	200 pips
J48	83,8%	99,9%	92,8%	97,6%	97,6%
NaiveBayes	59,9%	93,5%	87,2%	92,2%	92,2%
KStar	75,4%	98,1%	92,5%	97%	97%
MultilayerPerceptron	75,61%	97,6%	92,8%	97,4%	97,4%
ZeroR	64,4%	88%	87,7%	95,9%	95,9%

É presumível que os dados estão indo ao encontro da superadaptação nas tendências, pois, ao aumentar a quantidade do limiar a quantidade de exemplos de tendências para a construção do classificador é reduzida, assim o classificador não consegue generalizar suficientemente para realizar predições de fato assertivas, ou seja, apesar de um aumento da acurácia isso não reflete em um classificador que de fato atenda as necessidades de operar no mercado. Sendo inviável até então a construção de classificadores com limiares maiores de 50 pips.

Isto fica claro quando analisamos de perto a crescente acurácia do classificador de base ZeroR. Como ele é um classificador que utiliza a moda para classificar as demais classes, a classe principal se sobrepõe as demais, logo é possível concluir que com esta base de dados tem muitos exemplos para uma classe e as demais classes existe uma escassez de exemplos.

Ao realizar os testes utilizando os *datasets* harmonizados, com base no estudo apresentado na seção 7.1. Quando analisamos a acurácia dos classificadores realizando uma comparação entre as tabelas 7 e 8 é possível perceber até uma aproximação entre as porcentagens. Logo, o que diferencia, comprovando a viabilidade da utilização dos dados harmonizados combinado com um limiar menor, como apresentado na tabela 8 é quando olhamos o classificador de base ZeroR. Como dito anteriormente, ele é um classificador baseado em regras, a regra em questão classifica os elementos em relação a sua moda, sendo assim, quanto menor a acurácia deste classificador, maior será a quantidade de exemplos dados para as outras operações.

Tabela 8: Resultado das acurácias do experimento pragmático.

Classificador	acc 6 pips	acc 12 pips	acc 25 pips	acc 50 pips
J48	83,2%	84,6%	87,2%	90,3%
NaiveBayes	69,8%	76,6%	80,1%	83,9%
KStar	75,4%	85,2%	87,4%	90,3%
MultilayerPerceptron	82,6%	84%	86,6%	90,3%
ZeroR	51%	57%	68,1%	80,5%

Outro critério de desempate que podemos considerar é quando analisamos as matrizes de confusão. Como representado na tabela 9 é possível observar que quando construímos um classificador utilizando os *datasets* de forma harmonizada, os nossos exemplos da classe de compra e venda aumentam em 37,89% para os exemplos de compra e 66,3% para os exemplos de venda.

Tabela 9: Matriz de confusão do algoritmo ZeroR

Técnica de Classificação	Algoritmo	Comprar	Vender	Não Operar
Baseado em regras	ZeroR	a	b	c
Dados Não-Harmonizados	50 pips	0	0	664
		0	0	55
		0	0	5281
Baseado em regras	Zero R	a	b	c
Dados Harmonizados	50 pips	0	0	1069
		0	0	163
		0	0	5094

### 7.3.3 Concordância

Tal como feito na abordagem formal na seção 7.3.1.1. Foi calculado o coeficiente Kappa para cada limiar escolhido. A tabela 10 ilustra os resultados obtidos.

Observa-se que o classificador J48 obteve a melhor posição e o *NaiveBayes* continuou na última posição quando analisamos o coeficiente Kappa. Entretanto, quando

Tabela 10: Resultado dos coeficientes Kappa e sua interpretação.

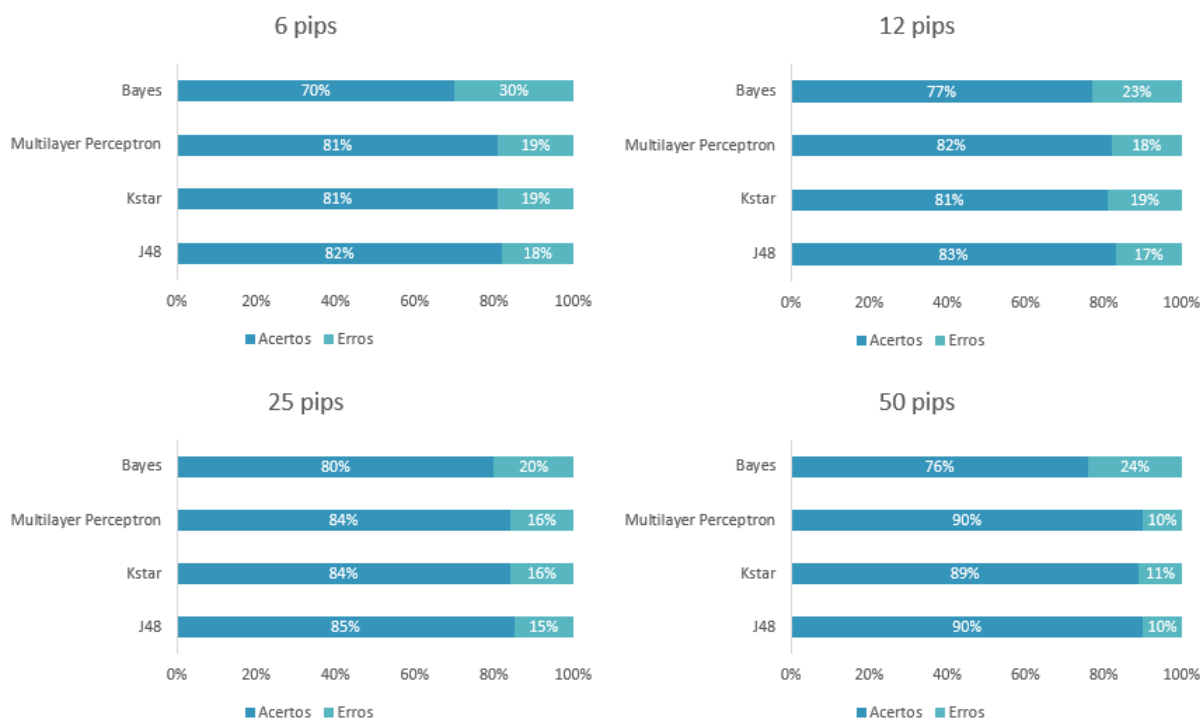
Classificador (Limiar)	Kappa	Interpretação
J48 (6 pips)	0,7011	Concordância Forte
NaiveBayes (6 pips)	0,5439	Concordância Moderada
KStar (6 pips)	0,7213	Concordância Forte
MultilayerPerceptron (6 pips)	0,6934	Concordância Forte
J48 (12 pips)	0,6997	Concordância Forte
NaiveBayes (12 pips)	0,5858	Concordância Moderada
KStar (12 pips)	0,7156	Concordância Forte
MultilayerPerceptron (12 pips)	0,6881	Concordância Forte
J48 (25 pips)	0,6939	Concordância Forte
NaiveBayes (25 pips)	0,5842	Concordância Moderada
KStar (25 pips)	0,6981	Concordância Forte
MultilayerPerceptron (25 pips)	0,6785	Concordância Forte
J48 (50 pips)	0,629	Concordância Forte Concordância Leve
NaiveBayes (50 pips)	0,5312	Concordância Moderada
KStar (50 pips)	0,637	Concordância Leve
MultilayerPerceptron (50 pips)	0,6452	Concordância Forte

utilizamos limiares de 50 pips, os classificadores J48 e KStar praticamente nenhuma concordância se manifestou. Isso talvez seja justificado por estes classificadores utilizarem o conceito de entropia para realizar o aprendizado, fundamentados em comportamentos determinísticos. Em contrapartida, os classificadores *NaiveBayes* e *Multilayer Perceptron*, mantiveram uma coerência na definição da concordância, possivelmente por se tratar de classificadores que detêm de comportamentos probabilísticos.

### 7.3.3.1 Previsões

Bem como feito na abordagem formal, também foram realizados os testes de previsões para a abordagem pragmática. A metodologia para a realização da previsão aqui foi a mesma, ou seja, a avaliação das previsões foi feita com base no cálculo de acerto entre o número de classe prevista sobre o número de classes no *dataset*

A figura 36 apresenta os cinco limiares escolhidos utilizados durante o estágio de teste. Observa-se que o classificador J48 manteve a sua colocação na primeira posição em todos os limiares, e analisando estatisticamente, todos os classificadores mantiveram suas porcentagens de forma equilibradas.

Figura 36: Percentagens do valores preditos utilizando os *datasets* de teste.

Fonte: Elaborado pelo autor.

## 7.4 Qualidade das previsões

Dada a complexidade nas séries temporais do mercado Forex fez-se necessário um último estudo de validação para comprovar a viabilidade do uso de classificadores em um ambiente real. Logo, foi verificado e contabilizado a ocorrência de atrasos nos resultados dos classificadores, pois, foi percebido que mesmo obtendo uma porcentagem média de 72% de acerto nas previsões, quando analisamos o resultado dos dados de teste foram percebidos atrasos na realização das operações.

Estes atrasos correspondem que as operações não foram iniciadas na primeira barra da tendência. A título de ilustração um atraso manifesta-se da seguinte forma; em uma tendência com 5 barras de *uptrend*, o classificador mandou operar a partir da terceira barra, quando o certo seria a operação iniciar desde a primeira barra, logo temos um atraso de 2 barras.

A tabela 11 apresenta ocorrência dos atrasos para a abordagem Formal, bem como a tabela 12 para a abordagem Pragmática. É possível notar que o atraso mínimo aparece em tendências que tenham no mínimo três barras e no máximo nove barras, ou seja, tendências que apresentam mais de nove barras não há a ocorrência de atrasos.

Também evidencia-se que os classificadores construídos com a abordagem formal apresenta uma ocorrência de atrasos bem superior aos classificadores construídos com a abordagem pragmática. A título de exemplo, quando comparamos o Classificador J48

Tabela 11: Contabilidade dos Atrasos - Abordagem Formal

Classificador: J48							
Atraso	3	4	5	6	7	8	9
Ocorrência (Quantidade)	11	6	0	0	0	0	0
Classificador: KStar							
Atraso	3	4	5	6	7	8	9
Ocorrência (Quantidade)	14	3	1	1	0	0	0
Classificador: Multilayer Perceptron							
Atraso	3	4	5	6	7	8	9
Ocorrência (Quantidade)	27	8	10	6	1	1	0
Classificador: NaiveBayes							
Atraso	3	4	5	6	7	8	9
Ocorrência (Quantidade)	2	4	3	1	2	0	0

da abordagem formal com o J48 da abordagem pragmática ocorre um aumento de mais de 1000% no número de casos para tendências com no mínimo 3 barras, bem como há aumento no número de casos quando observamos as demais tendências.

Tabela 12: Contabilidade dos Atrasos - Abordagem Pragmática

Classificador: J48							
Atraso	3	4	5	6	7	8	9
Ocorrência (Quantidade)	158	34	18	16	5	3	4
Classificador: KStar							
Atraso	3	4	5	6	7	8	9
Ocorrência (Quantidade)	139	26	17	17	8	4	4
Classificador: Multilayer Perceptron							
Atraso	3	4	5	6	7	8	9
Ocorrência (Quantidade)	148	26	18	12	2	3	2
Classificador: NaiveBayes							
Atraso	3	4	5	6	7	8	9
Ocorrência (Quantidade)	34	4	0	0	0	1	0

Um segundo fator que é possível estabelecer é um ranqueamento em relação ao classificador que obteve a menor ocorrência de atrasos. Como é possível observar nas tabelas supracitadas, o classificador em questão é o *NaiveBayes*, uma surpresa, já que dentre os demais classificadores em todas as abordagens este classificador apresentou desempenho inferior em comparação aos demais.

## 7.5 Análise estatística da Abordagem Híbrida.

De modo a comprovar a aplicação prática da teoria formal, foi realizado um estudo estatístico da abordagem híbrida que fora abordada na seção 28. A realização deste estudo justifica-se, pois, a abordagem formal caracteriza a formação das famílias de estado e a abordagem pragmática caracteriza a modelagem das tendências considerando os valores de cotação do mercado. Logo, é esta abordagem que une a teoria e a prática, ficando incumbida de construir um classificador confiável de fato. A figura 37 ilustra os resultados percentuais deste estudo.

Figura 37: Resultado percentual da análise estatística da abordagem híbrida.



Fonte: Elaborado pelo autor.

O primeiro gráfico de junção demonstra a quantidade de combinação entre as colunas de classe das duas abordagens (formal e pragmática). Nota-se que a medida que aumenta a quantidade do limiar, é diminuído a quantidade de exemplos para as outras classes. Em relação às classes que não tiveram combinação, não é definida nenhuma classe para estas instâncias, elas são retiradas do *datasets* para a construção do classificador. Sendo assim, resta apenas trabalhar com porcentagens limitadas para a construção do classificador, o que não traz uma confiabilidade dos resultados por conta do estudo de harmonização feito na seção 7.1

A figura também evidencia que quando olhamos a porcentagem de cada classe (Não Operar, Vender e Comprar) individualmente é possível analisar o percentual de ocorrência de cada família de estado relacionado à sua respectiva classe. Observa-se que as famílias

de estados que se relacionam com a classe não operar é comprovada. Uma atenção especial deve ser dada as famílias de estados que são reversões de tendência. Bom, é possível ver que a teoria de fato correlacionou-se com a prática. Sendo assim é possível comprovar a existência de um classificador otimizado que considere um modelo estruturado de forma teórica.

## 8 Conclusões

Este trabalho dedicou-se a investigar teórica e empiricamente a possibilidade de predição de séries temporais do mercado Forex. Por intermédio de uma metodologia proprietária, buscou-se modelar as séries temporais oriundas do mercado através do modelo de estados termodinâmicos, transformando a modelagem em questão em uma análise markoviana de segunda ordem. Esta análise teve como um dos objetivos a criação de uma base de dados compatível com as técnicas de mineração de dados para que pudesse ser construído e avaliado classificadores com a finalidade de prever tendências futuras no presente mercado. As avaliações foram realizadas através de análise estatística e comparação de resultados entre os classificadores preditivos, tais como, J48, KStar, Multilayer Perceptron e NaiaveBayes, cada um com suas nuances e especificidades. Tentando assim, comprovar a eficiência e adequação destes classificadores em relação à combinação de modelagem de séries temporais, buscando a predição e detecção de padrões no mercado financeiro.

Para verificar se a construção de classificadores são de fato viáveis para predições de tendências, foram feitos uma série de estudos, tais como; Estudo da harmonização da base de dados, análise estatística dos modelos de classificação, estudo das qualidades das predições e por fim o estudo estatístico da junção entre as duas abordagens propostas. Os resultados obtidos através dos estudos foram submetidos a critérios de avaliação com base em parâmetros estatísticos, tais como acurácia, matriz de confusão, coeficiente Kappa e análises comparativas empíricas.

Com os resultados preliminares foi possível fazer um ranqueamento entre os classificadores propostos, onde, mesmo com nuances diferentes, três entre os quatro classificadores avaliados tiveram desempenho superior a 70% em todos os estudos realizados, baseado no modelo proposto. Apesar de promissor, faz-se necessário apropriar-se do uso destas técnicas com cautela, por conta da estocasticidade das séries temporais do mercado Forex.

Por fim, foi possível constatar a transformação do modelo de estados termodinâmicos em forma de primeira ordem para segunda ordem. Adequação das técnicas de mineração de dados para uma base de dados supervisionada para o modelo proposto, este caráter dinâmico, contribui também para possíveis ajustes futuros, em razão de atualizações ou defasagem do modelo. Os resultados evidenciam que é possível obter padrões através da modelagem de séries temporais, caso estes padrões sejam interpretados por técnicas preditivas adequadas.

## 8.1 Trabalhos Futuros

A partir do estudo técnico-científico realizado no presente trabalho, sugere-se como ponto de partida a validação do modelo proposto com um volume maior de dados harmonizados (equilibrados) para que possa ser construída de fato o classificador da abordagem híbrida. Outra proposta é o refinamento do modelo markoviano de segunda ordem, no qual possa ser adicionado novos parâmetros do mercado ou realizar uma possível análise markoviana de terceira ordem, visto que esta transformação pode trazer novos *Insights* científicos.

Propõe-se também a criação de um sistema especialista que possa organizar de forma automatizada toda a metodologia, buscando integrar de forma eficiente os sub-estágios da mesma. Bem como à construção uma métrica estatística possa justificar todo o modelo proposto. Não obstante, é possível explorar novos classificadores e técnicas de mineração de dados bem como a adequação da metodologia proposta para outras áreas do conhecimento.

# Referências

- ACTION, P. *Classificação de Estados em uma Cadeia de Markov*. 2015. Disponível em: <<http://www.portalaction.com.br/processo-estocastico/52-classificacao-de-estados-em-uma-cadeia-de-markov>>. Citado 5 vezes nas páginas 29, 30, 31, 32 e 33.
- ADHIKARI, P. R. et al. Explaining mixture models through semantic pattern mining and banded matrix visualization. *Machine Learning*, v. 105, n. 1, p. 3–39, Oct 2016. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1007/s10994-016-5550-3>>. Citado na página 16.
- ADHIKARI, R.; AGRAWAL, R. An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*, 2013. Citado 4 vezes nas páginas 22, 23, 24 e 25.
- AGUIRRE, A. Uma introdução à análise espectral de séries temporais econômicas. *Nova Economia*, v. 5, n. 1, 2013. Citado na página 27.
- ALENCAR, A. B. *Mineração e visualização de coleções de séries temporais*. Tese (Doutorado) — Universidade de São Paulo, 2007. Citado 4 vezes nas páginas 23, 34, 36 e 37.
- AMO, S. D. Técnicas de mineração de dados. *Jornada de Atualização em Informática*, 2004. Citado 4 vezes nas páginas 34, 35, 36 e 37.
- ARORA, R. Comparative analysis of classification algorithms on different datasets using weka. *International Journal of Computer Applications*, Citeseer, v. 54, n. 13, 2012. Citado na página 54.
- ATTA, I. et al. Validação de nova teoria de estados e transições através de simulações com dados reais. *Universidade do Estado da Bahia (UNEB)*, p. 1–8, 2018. Citado 3 vezes nas páginas 15, 21 e 32.
- AVENT, S. R. *Time Series Analysis an Introduction*. 1999. Disponível em: <<http://userwww.sfsu.edu/efc/classes/biol710/timeseries/TimeSeriesAnalysis.html>>. Citado na página 24.
- BALLINI, R. et al. Análise e previsões de variações utilizando modelos de séries temporais, redes neurais e redes neurais nebulosas. [sn], 2000. Citado na página 14.
- BARRETO, G. et al. Modelagem computacional distribuída e paralela de sistemas e de séries temporais multivariáveis no espaço de estado. [sn], 2002. Citado na página 14.
- BARRETO, J. M. Introdução as redes neurais artificiais. *V Escola Regional de Informática. Sociedade Brasileira de Computação, Regional Sul, Santa Maria, Florianópolis, Maringá*, p. 5–10, 2002. Citado na página 26.
- BORTOLI, M. A. *Indicadores Baseados em Diferenças de Primeira Ordem para Orientar Robôs no Mercado FOREX*. 2015. UNEB (Universidade do Estado da Bahia), Salvador, Brasil. Citado 6 vezes nas páginas 19, 20, 22, 24, 26 e 62.

- BRESSAN, A. A. Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais. *RAE-eletrônica*, SciELO Brasil, v. 3, n. 1, 2004. Citado na página 14.
- BRIGGS, A.; SCULPHER, M. An introduction to markov modelling for economic evaluation. *PharmacoEconomics*, v. 13, n. 4, p. 397–409, Apr 1998. ISSN 1179-2027. Disponível em: <<https://doi.org/10.2165/00019053-199813040-00003>>. Citado na página 16.
- CAMILO, C. O.; SILVA, J. C. d. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. *Universidade Federal de Goiás (UFG)*, p. 1–29, 2009. Citado 2 vezes nas páginas 16 e 38.
- CARRIJO, I. B. *Extração de regras operacionais ótimas de sistemas de distribuição de água através de algoritmos genéticos multiobjetivo e aprendizado de máquina*. Tese (Doutorado) — Universidade de São Paulo, 2004. Citado na página 64.
- CARVALHO, J. L. d. S. d. *Predição da Tendência de Séries Temporais do Mercado FOREX Utilizando Indicador Baseado em Estatística Descritiva e Redes Neurais Recorrentes*. 2017. UNEB (Universidade do Estado da Bahia), Salvador, Brasil. Citado 6 vezes nas páginas 21, 25, 26, 27, 40 e 42.
- GONÇALVES, E. C. *Extração de Árvores de Decisão com a Ferramenta de Data Mining Weka*. 2007. Disponível em: <<https://www.devmedia.com.br/extracao-de-arvores-de-decisao-com-a-ferramenta-de-data-mining-weka/3388>>. Citado na página 51.
- GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. Previsão de desempenho de estudantes em cursos ead utilizando mineração de dados: uma estratégia baseada em séries temporais. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2012. v. 23, n. 1. Citado na página 14.
- HINOJOSA, F. R. A arte da pesquisa. *Educação: Teoria e Prática*, v. 20, n. 34, p. 189, 2010. Citado na página 41.
- HIRA, S.; DESHPANDE, P. S. Mining precise cause and effect rules in large time series data of socio-economic indicators. *SpringerPlus*, v. 5, n. 1, p. 1625, Sep 2016. ISSN 2193-1801. Disponível em: <<https://doi.org/10.1186/s40064-016-3292-0>>. Citado na página 14.
- IEZZI, G.; MURAKAMI, C. *Fundamentos de matemática elementar, 1: conjuntos, funções*. [S.l.]: Atual, 1993. Citado na página 24.
- INFOMONEY. *Como funciona o mercado de ações*. 2014. Disponível em: <<http://www.infomoney.com.br/infograficos/mercado-de-acoes>>. Citado na página 18.
- ISHII, R. P.; RIOS, R. A.; MELLO, R. F. Classification of time series generation processes using experimental tools: a survey and proposal of an automatic and systematic approach. *International Journal of Computational Science and Engineering*, Inderscience Publishers, v. 6, n. 4, p. 217–237, 2011. Citado 2 vezes nas páginas 22 e 23.

KEIEL, G.; BENDER, F. A. Modelagem de séries temporais financeiras: uma abordagem estatística para a identificação de modelos de média condicional. *Scientia cum Industria*, v. 6, n. 1, p. 22–28, 2018. Citado 2 vezes nas páginas 21 e 25.

KOSHIYAMA, A. S. et al. Modelagem de séries temporais aplicadas no estudo da taxa de câmbio brasileira. 2018. Citado 2 vezes nas páginas 21 e 25.

LATORRE, M. d. R. D. d.; CARDOSO, M. R. A. et al. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 4, p. 145–152, 2001. Citado 2 vezes nas páginas 14 e 21.

MANHÃES, L. M. B. et al. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2011. v. 1, n. 1. Citado na página 51.

MARQUES, A. F. d. C. *Desafios na previsão de séries temporais financeiras: o caso da taxa de câmbio EUR/USD*. Tese (Doutorado), 2017. Citado na página 14.

MARQUES, J. M.; SILVA, S. d. C. R. da. A cadeia de markov na determinação de indicadores educacionais. *Revista da FAE*, v. 16, n. 2, p. 88–101, 2013. Citado 4 vezes nas páginas 15, 28, 30 e 31.

MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: *Análise de séries temporais*. [S.l.: s.n.], 2006. Citado 3 vezes nas páginas 21, 24 e 25.

MORETTIN, P. A.; TOLOI, C. M. de C. *Modelos para previsão de séries temporais*. [S.l.]: Instituto de matemática pura e aplicada, 1981. Citado 3 vezes nas páginas 25, 26 e 27.

NÉRI, E. L. Agentes de software: delegando decisões a programas. *RAE-eletrônica*, SciELO Brasil, v. 4, n. 1, 2005. Citado na página 41.

NERY, A. R. L.; BASSI, A. B. M. S. et al. Condições de equilíbrio termodinâmico: a função disponibilidade. *Química Nova*, Sociedade Brasileira de Química, 2011. Citado na página 15.

NOGUEIRA, F. Cadeias de markov. *Universidade Federal de Juiz de Fora*, Universidade Federal de Juiz de Fora, v. 1, n. 1, 2009. Citado 2 vezes nas páginas 30 e 31.

OLIVEIRA, M. A.; FAVERO, L. P. L. Uma breve descrição de algumas técnicas para análise de séries temporais: Séries de fourier, wavelets, arima, modelos estruturais para séries de tempos e redes neurais. *VI SEMEAD Ensaio mqi. USP. Anais, São Paulo*, 2002. Citado na página 27.

OLIVEIRA, M. A. d. *Aplicação de redes neurais artificiais na análise de séries temporais econômico-financeiras*. Tese (Doutorado) — Universidade de São Paulo, 2007. Citado na página 22.

PRADHAN, G. N.; PRABHAKARAN, B. Association rule mining in multiple, multidimensional time series medical data. *Journal of Healthcare Informatics Research*, v. 1, n. 1, p. 92–118, Jun 2017. ISSN 2509-498X. Disponível em: <<https://doi.org/10.1007/s41666-017-0001-x>>. Citado na página 14.

- RODRIGUES, C. T. Marcio de C. Análise do desempenho da árvore de decisão para classificação preditiva de clientes em estado de recuperação judicial. *Congresso UnB de Contabilidade e Governança*, Universidade de Brasília, v. 1, n. 1, 2007. Citado 2 vezes nas páginas 52 e 54.
- ROLEDENE, S. et al. Genibux - event based intelligent forex trading strategy enhancer. *IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, p. 1–6, 2016. Citado 2 vezes nas páginas 15 e 18.
- ROSSI, P. O mercado internacional de moedas, o carry trade e as taxas de câmbio. *Observatório da Economia Global*, n. 5, 2010. Citado 2 vezes nas páginas 18 e 19.
- SAHOO, G.; KUMAR, Y. Analysis of parametric & non parametric classifiers for classification technique using weka. *International Journal of Information Technology and Computer Science (IJITCS)*, v. 4, n. 7, p. 43–49, 2012. Citado na página 54.
- SANTOS, M. M. S. R. dos. *Proposição de Modelo Matemático para Estratégia de Hedging baseada em Contraordem*. 2016. UNEB (Universidade do Estado da Bahia), Salvador, Brasil. Citado na página 26.
- SANTOS, R. J. Cadeias de markov. *UFMG/Departamento de Matemática. Disponível em: <http://www.mat.ufmg.br/~regi>*. Acesso em, v. 15, n. 10, 2008. Citado na página 29.
- SATO, R. C.; ZOUAIN, D. M. Modelos de markov aplicados a saúde. *Einstein*, v. 8, n. 3, p. 376–79, 2010. Citado na página 28.
- SIGNORETTI, R. *Classification of Markov chains states*. 2013. Disponível em: <<https://www.youtube.com/watch?v=qfxTVf1myCg>>. Citado 3 vezes nas páginas 29, 30 e 31.
- SILVA, M. A. P.; FILHO, E. A. S. da S. Sistema fuzzy tipo 2 para operação ao day-trade no mercado forex. *Jornada Atualização em Informática*, 2018. Citado 2 vezes nas páginas 19 e 20.
- SOUTO, M. de. Weka. *Centro de Informática UFPE*, Centro de Informática da Universidade Federal de Pernambuco., v. 1, n. 1, 2004. Citado na página 51.
- SOUZA, P. Uma estratégia baseada em algoritmos de mineração de dados para validar plano de operação de voo a partir de previsões de estados dos satélites do inpe. *São José dos Campos, SP, Brasil: INPE*, 2011. Citado 8 vezes nas páginas 34, 36, 37, 38, 39, 54, 55 e 66.
- SOUZA, R. M. d. Meta-aprendizado para seleção automática de modelos de séries temporais. Universidade Federal de Pernambuco, 2010. Citado na página 53.
- TEOFILO, D. *KDD – Knowledge Discovery in Database*. 2015. Disponível em: <<https://danielteofilo.wordpress.com/2015/02/16/kdd-knowledge-discovery-in-database/>>. Citado na página 35.
- VALOR. *Volume médio diário na Bovespa sobe em julho*. 2016. Disponível em: <<http://www.valor.com.br/financas/4658775/volume-medio-diario-na-bovespa-sobe-em-julho>>. Citado na página 18.