



CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

PREDIÇÃO DE TENDÊNCIAS EM SÉRIES FINANCEIRAS UTILIZANDO META-CLASSIFICADORES

CARLOS ALBERTO SILVA DE ASSIS

Orientador: Adriano C. Machado Pereira
CEFETMG-PPGMMC | UFMG-PPGCC

Coorientador: Eduardo Gontijo Carrano
UFMG-PPGEE

BELO HORIZONTE
ABRIL DE 2019

CARLOS ALBERTO SILVA DE ASSIS

**PREDIÇÃO DE TENDÊNCIAS EM SÉRIES
FINANCEIRAS UTILIZANDO
META-CLASSIFICADORES**

Tese apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Métodos Matemáticos Aplicados

Orientador: Adriano C. Machado Pereira
CEFETMG-PPGMMC | UFMG-PPGCC

Coorientador: Eduardo Gontijo Carrano
UFMG-PPGEE

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL
BELO HORIZONTE
ABRIL DE 2019

A848p Assis, Carlos Alberto Silva de
Predição de tendências em séries financeiras utilizando meta-
classificadores / Carlos Alberto Silva de Assis. – 2019.
90 f.

Tese de doutorado apresentada ao Programa de Pós-Graduação em
Modelagem Matemática e Computacional.

Orientador: Adriano C. Machado Pereira.

Coorientador: Eduardo Gontijo Carrano.

Tese (doutorado) – Centro Federal de Educação Tecnológica de
Minas Gerais.

1. Inteligência artificial – Teses. 2. Finanças – Modelos matemáticos
– Teses. 3. Matemática financeira – Teses. I. Pereira, Adriano Cesar
Machado. II. Carrano, Eduardo Gontijo. III. Centro Federal de Educação
Tecnológica de Minas Gerais. III. Título.

CDD 006.3



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

**PREDIÇÃO DE TENDÊNCIAS EM SÉRIES FINANCEIRAS
UTILIZANDO META-CLASSIFICADORES.**

Tese de Doutorado apresentada por **Carlos Alberto Silva de Assis**, em 24 de abril de 2019, ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG, e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Adriano César Machado Pereira
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Eduardo Gontijo Carrano
Universidade Federal de Minas Gerais

Prof. Dr. Marconi de Arruda Pereira
Universidade Federal São João del-Rei

Prof. Dr. Felipe Dias Paiva
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Daniel Hasan Dalip
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Flávio Luis Cardeal Pádua
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Rodrigo Tomás Nogueira Cardoso
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida à impressão,

Prof. Dr. Thiago de Souza Rodrigues
Coordenador do Programa de Pós-Graduação Stricto Sensu em
Modelagem Matemática e Computacional

Dedico meu trabalho aos meus filhos Lucas Assis e Guilherme Viana, minha mãe e em especial, a minha esposa Karine Assis e minha afilhada Ana Carolina Alcântara.

Agradecimentos

Agradeço primeiramente a Deus pela vida.

Agradeço a todos os colegas do grupo de Finanças Computacionais (FICO) pelo apoio e participação direta nesta pesquisa, em especial ao grande amigo e parceiro de curso, Eduardo Jabbur.

Destaco a total colaboração, incentivo, paciência e aprendizado oferecido pelos meus orientadores, Dr. Adriano César Machado Pereira e Dr. Eduardo Gontijo Carrano.

Agradeço aos professores participantes da banca examinadora neste momento tão importante e esperado: Dr. Rodrigo Tomás Nogueira Cardoso, Dr. Felipe Dias Paiva, Dr. Marconi de Arruda Pereira, Dr. Flávio Luis Cardeal Pádua e Dr. Daniel Hasan Dalip.

Agradeço ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, pela oportunidade e privilégio de realizar este trabalho, e sobretudo pelo aprendizado. Estendo esse agradecimento ao corpo docente e discente, com quem tanto aprendi.

À CAPES pelo apoio financeiro parcial no período de realização desta pesquisa.

“Sem desandar, sem humilhar ninguém. É assim que eu quero ser, sim, um cara melhor. Não melhor do que ninguém, mas o melhor que eu puder ser!” (Alexandre Magno Abrão - Chorão Charlie Brown Jr.)

Resumo

A previsão do comportamento de ativos financeiros é uma linha de pesquisa que vem sendo investigada por diversas técnicas ao longo dos últimos anos. Mesmo com inúmeras pesquisas, prever preços de ativos ou tendências continua sendo uma tarefa extremamente difícil, uma vez que tal comportamento está ligado às incertezas do mercado financeiro e outros fatores. Desta forma, neste trabalho foi desenvolvido um meta-classificador baseado em métodos de inteligência computacional para descobrir tendências de preço para ativos de bolsa de valores, como a B3. O *kernel* do meta-classificador é baseado na ferramenta *WEKA*, onde 7 classificadores são combinados para serem otimizados na etapa seguinte pela meta-classificação. Testes foram realizados com alguns dos ativos mais líquidos de diferentes setores e o ativo que acompanha o índice Bovespa da B3, são eles: BOVA11, CIEL3, ITUB4, PETR4, USIM5, CMIG4, GGBR4, KROT3 e GOLL4. Os resultados foram satisfatórios, apresentando uma boa acurácia na classificação com até 57%, além de resultados financeiros com ganhos de até 100% do valor de capital inicialmente investido. Também tivemos bons resultados quando comparamos com os *baselines buy-and-hold*, aleatório e estratégia inversa.

Palavras-chave: Inteligência Computacional. Meta-Classificador. Séries Financeiras.

Abstract

Predicting the behavior of financial assets is a task that has been researched by various techniques over the last years. Despite there exists an extensive research in this area, the task to predict asset prices or trends remains an extremely difficult task because due to the uncertainties of the financial markets and other factors. This work proposes and implement a meta-classifier based on computational intelligence techniques to find price trends for the stock market assets, as the B3. Meta-classifier kernel is based on the WEKA tool, where 7 classifiers are combined to be optimized in the next step by meta-classification. Tests were performed with some of the most liquidity assets of different sectors and the assets that accompany the Bovespa index of B3, are: BOVA11, CIEL3, ITUB4, PETR4, USIM5, CMIG4, GGBR4, KROT3 and GOLL4. The results were satisfactory, showing a good accuracy in the classification with up to 57%, in addition to financial results with gains of up to 100% of the capital value initially invested. We also had good results when compared to the buy-and-hold, random and inverse strategy.

Keywords: Computational Intelligence. Meta-Classifier. Financial Series.

Lista de Figuras

Figura 1 – Exemplo de Série Temporal Financeira	14
Figura 2 – <i>Candlestick</i>	19
Figura 3 – Gráfico de <i>Candlesticks</i>	20
Figura 4 – Processo de classificação.	23
Figura 5 – Classificação de um conjunto de dados usando uma SVM linear.	23
Figura 6 – Modelo computacional de um Neurônio Artificial.	25
Figura 7 – Exemplo de uma Rede Neural Artificial Multilayer Perceptron (MLP).	26
Figura 8 – Exemplo de uma saída de Árvore de Decisão (J48).	27
Figura 9 – Grafo G com ordem e tamanho igual a 5.	28
Figura 10 – Fluxograma do funcionamento básico de um Programa Genético.	30
Figura 11 – Exemplo de indivíduo na Programação Genética.	31
Figura 12 – Exemplo de cruzamento na Programação Genética.	34
Figura 13 – Exemplo de mutação na Programação Genética.	35
Figura 14 – Representação do algoritmo <i>Stacking</i>	37
Figura 15 – Matriz de Confusão.	38
Figura 16 – Metodologia Proposta.	41
Figura 17 – Exemplo de Janela Deslizante.	43
Figura 18 – Série de Preços de Fechamento Original.	52
Figura 19 – Distribuição das classes por ativo.	53
Figura 20 – Comparação de Dispersão Geral dos Retornos Financeiros.	61
Figura 21 – Principais Etapas da Revisão Sistemática da Literatura.	78
Figura 22 – Comparação Drawdown e % Retornos Financeiros - BOVA11.	82
Figura 23 – Comparação Drawdown e % Retornos Financeiros - CIEL3.	83
Figura 24 – Comparação Drawdown e % Retornos Financeiros - ITUB4.	84
Figura 25 – Comparação Drawdown e % Retornos Financeiros - PETR4.	85
Figura 26 – Comparação Drawdown e % Retornos Financeiros - USIM5.	86
Figura 27 – Comparação Drawdown e % Retornos Financeiros - CMIG4.	87
Figura 28 – Comparação Drawdown e % Retornos Financeiros - GGBR4.	88
Figura 29 – Comparação Drawdown e % Retornos Financeiros - KROT3.	89
Figura 30 – Comparação Drawdown e % Retornos Financeiros - GOLL4.	90

Lista de Tabelas

Tabela 1 – Valor de Mercado de Ativos entre Agosto de 2015 e Agosto de 2016. . .	52
Tabela 2 – <i>Benchmarks</i>	52
Tabela 3 – Parâmetros SVM.	53
Tabela 4 – Parâmetros RF.	53
Tabela 5 – Parâmetros MLP.	54
Tabela 6 – Parâmetros J48.	54
Tabela 7 – Parâmetros SMO.	54
Tabela 8 – Parâmetros PG.	54
Tabela 9 – Intervalo de parâmetros de calibração do Meta-Classificador	55
Tabela 10 – Comparação de Retorno Financeiro Acumulado entre pares de Classificadores e Meta-Classificadores.	57
Tabela 11 – Comparação da Acurácia entre pares de Classificadores e Meta-Classificadores.	58
Tabela 12 – Comparação de Retorno Financeiro entre melhor Classificador com Meta-GP.	58
Tabela 13 – Comparação de Retorno Financeiro entre melhor Classificador com Meta-Classificador.	58
Tabela 14 – Exemplos de Custos Operacionais.	59
Tabela 15 – Referência de <i>Drawdown</i>	61
Tabela 16 – Resultados Meta-Classificador - Meta-RF	62
Tabela 17 – Resultados Classificador - MLP.	62
Tabela 18 – Resultado de % Retorno Financeiro dos <i>Baselines</i>	63
Tabela 19 – Resultado RSL: Quantidade de trabalhos selecionados por área pesquisada.	81

Lista de Abreviaturas e Siglas

ANN	Rede Neural Artificial
ARMA	Modelo Auto-regressivo de Médias Móveis.
BD	Base de Dados
BN	Redes Bayesianas
CB	Critérios de Busca
CDB	Certificado de Depósito Bancário
CE	Critérios de Exclusão
CI	Critérios de Inclusão
CNN	Redes Neurais Covolutivas
CQ	Critérios de Qualidade
DC	Dados Coletados
FB	Ferramentas de Busca
FN	Falsos Negativo
FP	Falsos Positivo
GARCH	Modelo Autoregressivo de Heteroscedasticidade Condicional Generalizado
IC	Inteligência Computacional
J48	Árvore de Decisão
KNN	Vizinho Mais Próximo
LSTM	Redes Neurais Recorrentes
MLP	Perceptron Multi-Camadas
MQL	MetaQuotes Language
PG	Programação Genética
PQ	Programação Quadrática

QP	Questões de Pesquisa
RBM	Máquinas Restritas de Boltzmann
RF	Florestas Aleatórias
RNA	Redes Neurais Artificiais
RSL	Revisão Sistemática da Literatura
SELIC	Sistema Especial de Liquidação e de Custódia
SMO	Otimização Mínima Sequencial
SVM	Máquinas de Vetor de Suporte
SVR	Regressão Vetorial de Suporte
VN	Verdadeiros Negativo
VP	Verdadeiros Positivo

Sumário

1 – Introdução	1
1.1 Definição do Problema de Pesquisa	2
1.2 Objetivos geral e específicos	3
1.3 Contribuições	4
1.4 Organização do trabalho	5
2 – Revisão da Literatura	6
2.1 Trabalhos Relacionados	6
2.1.1 Inteligência Computacional em Séries Temporais Financeiras	7
2.1.2 Meta-Classificadores em Geral	10
2.1.3 Meta-Classificadores para Mercado Financeiro	10
2.1.4 Considerações Finais	12
3 – Fundamentação Teórica	13
3.1 Séries Temporais Financeiras	13
3.2 Mercado Financeiro	15
3.2.1 Análise Técnica	17
3.2.2 Estratégias de Negociação	19
3.3 Inteligência Computacional, Ferramentas e Algoritmos	21
3.3.1 <i>Waikato Environment for Knowledge Analysis (WEKA)</i>	22
3.3.2 Classificação	22
3.3.3 Máquinas de Vetores de Suporte (SVM)	22
3.3.4 Redes Neurais Artificiais (RNA)	25
3.3.5 Árvore de Decisão (J48)	26
3.3.6 Otimização Mínima Sequencial (SMO)	27
3.3.7 Florestas Aleatórias (RF)	28
3.3.8 Redes Bayesianas (BN)	28
3.3.9 Programação Genética (PG)	29
3.3.9.1 Indivíduo	31
3.3.9.2 Operadores Genéticos	32
3.3.9.3 Seleção	32
3.3.9.4 Cruzamento	33
3.3.9.5 Mutação	33
3.3.10 Multi-Classificadores	34
3.3.10.1 Meta-Classificadores	36
3.4 Métricas de Avaliação	37

3.4.1	Medidas de Desempenho	37
3.4.2	Medidas Financeiras	39
3.5	Considerações Finais	40
4	– Metodologia	41
4.1	Extração dos Dados	41
4.2	Tratamento dos Dados	42
4.3	Meta-classificação	43
4.4	Simulação do Mercado	45
4.5	Análise dos Resultados	47
4.6	Considerações Finais	49
5	– Experimentos e Resultados	50
5.1	Caracterização e Análise dos Dados	51
5.2	Instanciação dos Algoritmos	53
5.3	Avaliação dos Modelos e Estratégias	55
5.4	Avaliação da Viabilidade dos Custos Operacionais	57
5.5	Resultados Experimentais	59
5.5.1	Considerações Finais	61
6	– Conclusão	64
6.1	Trabalhos futuros	65
	Referências	67
	Apêndices	76
	APÊNDICE A – Revisão Sistemática da Literatura (RSL)	77
A.0.1	Planejamento	77
A.0.2	Execução	80
A.0.3	Análise dos Resultados	81
	APÊNDICE B – BOVA11	82
	APÊNDICE C – CIEL3	83
	APÊNDICE D – ITUB4	84
	APÊNDICE E – PETR4	85
	APÊNDICE F – USIM5	86

APÊNDICE G – CMIG4	87
APÊNDICE H – GGBR4	88
APÊNDICE I – KROT3	89
APÊNDICE J – GOLL4	90

Capítulo 1

Introdução

O mercado de ações é uma parte importante de qualquer economia e, por isso, seu entendimento é constante alvo de vários estudos. Investidores, economistas e pesquisadores de diversas áreas procuram modelos e informações que possam ajudá-los a entender e, possivelmente, prever o comportamento do mercado.

A previsão do comportamento da série de retornos de ativos do mercado financeiro sempre foi alvo de grande interesse em especial de investidores, buscando maneiras de maximizar ganhos financeiros através da identificação de oportunidades na compra e venda de ativos.

Adicionalmente, com as mudanças observadas na economia brasileira nos últimos anos, o comportamento do investidor brasileiro vem mudando progressivamente. Se antes era, de certa forma, vantajoso guardar dinheiro na caderneta de poupança, hoje em dia isto já não é mais verdade. Com a queda dos juros da SELIC, os rendimentos fornecidos pela poupança estão muito baixos. O mesmo acontece com outras formas de investimento seguras, ou melhor, que oferecem baixo risco de perda de capital ao investidor, como a renda fixa e os certificados de depósito bancário (CDBs). Atualmente, o investidor que deseja obter uma lucratividade maior precisa migrar para formas de investimento mais arriscadas, que assim são definidas porque, apesar de oferecerem a possibilidade de lucros maiores, podem também trazer ao investidor prejuízos maiores, se mal utilizadas.

A verdade é que não há uma fórmula mágica para prever o mercado financeiro ([BARRY-MORE, 2017](#)). Diante da grande dificuldade desta previsão, o mercado de ações demanda a criação de diversas técnicas que auxiliem em cenários para a tomada de decisões ([INFOMONEY, 2017](#)).

Sendo o mercado de ações um ambiente de grandes oportunidades, existem diversos estudos sobre este tema já desenvolvidos e outros tantos em construção ([TKáč; VERNER, 2016](#); [CAVALCANTE et al., 2016](#)). Na maioria desses estudos, o grande desafio é minimizar o erro de previsão, auxiliando na tarefa de prever o *timing* do mercado, isto é, o melhor

momento de compra ou venda de um ativo, maximizando o lucro. Para isso, são utilizados diversos modelos aplicados aos dados históricos a fim de prever seu comportamento futuro. Esses dados podem ser vistos como uma série temporal¹.

A previsão de séries temporais financeiras é uma das questões mais pesquisadas no campo das finanças, sobretudo no que diz respeito ao mercado acionário; prever o seu comportamento futuro não é uma tarefa simples (LIMA, 2017). Para tanto, essas pesquisas envolvem desde modelos estatísticos e econométricos até modelos de inteligência computacional².

Alguns exemplos da utilidade das previsões financeiras são:

- Gerar modelos de análise de crédito para caracterizar clientes que pagam em dia, clientes que pagam em atraso e clientes que não pagam;
- Desenvolver mecanismos de detecção de fraudes em compras de cartão de crédito;
- Ainda na área de finanças, construir e avaliar modelos de predição de séries temporais a partir do histórico de cotações de ações na bolsa de valores.

Enfim, muitas são as formas de se utilizar a previsão em séries temporais financeiras de modo a obter melhores resultados em problemas reais.

1.1 Definição do Problema de Pesquisa

Embora os erros de previsão sejam inevitáveis, diversas técnicas exploram modelos na busca de minimizar os erros de predição, utilizando um arsenal de métodos estatísticos e de inteligência computacional.

O problema de pesquisa abordado nesta tese é um tema interdisciplinar que envolve diversas áreas do conhecimento como, por exemplo, finanças, matemática, estatística, economia, ciência de dados e aprendizado de máquina de forma integrada e aplicada. Diante deste cenário, o problema de pesquisa foi decomposto em quatro questões principais a serem investigadas:

- **Tratamento dos Dados:** devido à grande quantidade de fontes de dados, existe a necessidade de entender melhor o comportamento e a relevância de alguns ativos, setores, indicadores e índices de mercado. Deste modo, pretendemos responder a seguinte questão: Qual a melhor forma e/ou técnicas de ciências dos dados para selecionar, preparar e analisar os dados de séries de cotação histórica sobre o

¹Uma série temporal é um conjunto de observações de um fenômeno ordenadas no tempo (BROCKWELL; DAVIS, 2002)

²Inteligência Computacional é uma área da Ciência que tem como objetivo desenvolver e aplicar recursos que permitam aos computadores terem um comportamento similar ao humano em tarefas específicas (GOLDSCHMIDT, 2010).

mercado financeiro?

- **Modelo de Previsão:** na literatura existem inúmeros algoritmos de classificação desenvolvidos, os quais consistem na generalização de experiências passadas para prever algum valor discreto (conhecido como classe) em novos exemplos (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Com base neste problema, definiu-se a seguinte questão: Quais algoritmos escolher? Outro tema importante é que existem estudos que mostram que combinar classificadores geralmente traz melhores resultados que os classificadores isolados. Este tipo de técnica considera, por exemplo, a estratégia utilizada quando uma pessoa precisa tomar uma decisão difícil. Normalmente, prefere-se considerar opiniões de vários especialistas em vez de considerar a opinião de apenas um. Assim, complementa-se a questão: é melhor combinar algoritmos no problema em questão? Quais técnicas utilizar?
- **Parametrização dos Modelos de Previsão:** cada algoritmo selecionado depende de inúmeras configurações, dificultando diretamente a qualidade do aprendizado do modelo de previsão. Com base neste problema, surge a seguinte questão: como configurar, treinar e medir tais algoritmos de classificação para a geração de previsão de tendências que garantam uma taxa mínima de aprendizado em relação à sua precisão como medida de desempenho?
- **Modelo de Operação:** é um grande desafio modelar estratégias de operação para o mercado financeiro a fim de identificar quais delas são melhores ou mais adequadas para diferentes objetivos almejados em relação ao risco/retorno. Portanto, outra questão de pesquisa seria: De que forma podemos aplicar uma série de previsão de tendências como estratégia de operação buscando resultados cujo percentual de retorno financeiro seja viável, rentável e que supere os valores de referência dos *baselines* para a aplicação em um ambiente real do mercado financeiro?

1.2 Objetivos geral e específicos

O objetivo geral deste trabalho é apresentar um método para a previsão de séries temporais financeiras utilizando um conjunto de técnicas de inteligência computacional combinadas. Pretende-se usar as técnicas de Programação Genética (PG), Máquinas de Vetor de Suporte (SVM), Florestas Aleatórias (RF), Perceptron Multicamadas (MLP), Árvore de Decisão (J48), Otimização Mínima Sequencial (SMO) e Redes Bayesianas (BN) a fim de criar um meta-classificador com o intuito de minimizar os erros de previsão.

O que se pretende é utilizar as saídas de um conjunto de classificadores base para o treinamento de um outro conjunto de meta-classificadores para, então, fornecer a predição final. A explicação para o uso dessa técnica é o fato de que, segundo Wolpert (1992), ao passar pelas camadas, o meta-classificador aprende com os erros dos classificadores

anteriores.

O método proposto recebe como entrada os valores dos períodos passados e retorna uma previsão que indica o que irá acontecer com a série temporal financeira sendo analisada no próximo período: se irá subir ou não. Com esse resultado, um algoritmo de estratégia pode decidir o que fazer: vender ou comprar a ação. Os resultados são apresentados por meio do índice de acertos do método, a variação de capital que um investidor hipotético teria obtido se tivesse seguido todas as operações recomendadas, e uma comparação desta variação com a que seria obtida com outras abordagens comparativas, como: *buy-and-hold*, aleatório e estratégia inversa (JOHNSTON, 2003). Para validar o método em diferentes situações, avaliamos seu comportamento com um conjunto de ações do mercado brasileiro com granularidade diária.

Desta forma, espera-se obter retornos financeiros acima dos ganhos da média do mercado financeiro. Para alcançar esta meta, os objetivos específicos são:

- Revisar a literatura referente a métodos de previsão de séries temporais financeiras, especialmente trabalhos que envolvem técnicas de inteligência computacional.
- Construir o arcabouço de meta-classificação, por meio de ferramentas tecnológicas adequadas;
- Realizar testes utilizando dados de séries reais financeiras da Bolsa de Valores de São Paulo (B3);
- Apresentar os resultados de métricas de performance;
- Apresentar os resultados de estratégias financeiras simuladas;

1.3 Contribuições

A principal contribuição deste trabalho é a construção de um arcabouço que implementa a abordagem do previsor baseado em meta-classificadores e estratégias de investimento.

Assim, durante todo o período de doutorado, inúmeras iniciativas, construções, tentativas, combinações e avaliações foram realizadas com o intuito de resolver os problemas em questão e, como consequência deste trabalho realizado, os tópicos discutidos nesta tese geraram as produções bibliográficas descritas a seguir.

Em congressos nacionais e internacionais:

- Carlos Assis, Adriano Pereira, Eduardo Carrano. “**Forecasting Financial Time Series using Meta-Classifier based on Genetic Programming**”. Econophysics Colloquium. 2016. Status: Publicado. Qualis: N/A.
- Carlos Assis, Adriano Pereira, Eduardo Carrano, Rafael Ramos, Wanderson Dias.

“Restricted Boltzmann Machines for the Prediction of Trends in Financial Time Series”. International Joint Conference on Neural Networks (IJCNN). 2018. Status: Publicado. Qualis: A1 - Ciência da Computação.

Em periódicos nacionais e internacionais:

- Carlos Assis, Eduardo Machado, Adriano Pereira, Eduardo Carrano. **“Hybrid deep learning approach for financial time series classification”**. Revista Brasileira de Computação Aplicada. 2018. Status: Publicado. Qualis: B2 - Interdisciplinar.

1.4 Organização do trabalho

Este trabalho está organizado em seis capítulos. Neste capítulo, introduzimos o tema, apresentando os objetivos e, posteriormente, foram detalhadas as motivações justificando o estudo do problema de previsão de séries temporais financeiras, além das contribuições do trabalho. No capítulo 2 apresentamos uma metodologia de revisão bibliográfica chamada Revisão Sistemática da Literatura (RSL), juntamente com a revisão de algumas das principais técnicas de séries temporais financeiras encontradas na literatura. No capítulo 3 mostraremos todo o referencial teórico utilizado como base para a formalização deste trabalho. Serão apresentados todos os algoritmos e conceitos necessários para a construção da solução proposta. No capítulo 4 será exposta toda a metodologia, serão descritas todas as ferramentas utilizadas, os conjuntos de dados históricos, os métodos de transformação dos dados, a forma de validação dos resultados, em resumo, todo o arcabouço desenvolvido. No capítulo 5 será discutida a análise dos resultados obtidos; apresentaremos resultados com métricas de performance e resultados financeiros simulados. No capítulo 6, será apresentada uma conclusão com uma análise das principais contribuições geradas. Além disso, serão apresentados os próximos passos da pesquisa. Por fim, os apêndices B, C, D, E, F, G, H, I e J apresentam os resultados gráficos de comparação gerados.

Capítulo 2

Revisão da Literatura

Este capítulo tem como objetivo identificar e analisar trabalhos existentes relacionados ao tema proposto. Para realizar este trabalho de forma metódica, realizou-se uma Revisão Sistemática da Literatura (RSL) proposta por [Kitchenham \(2004\)](#).

No apêndice [A](#) apresenta de forma geral os principais conceitos e diretrizes desta metodologia, e sua aplicação neste trabalho.

2.1 Trabalhos Relacionados

Nesta Seção serão apresentados os principais trabalhos relacionados ao tema que foram encontrados após a aplicação da RSL.

De modo geral, as técnicas de previsão do mercado financeiro podem ser classificadas em três categorias, como segue:

- **Métodos convencionais:** são aqueles criados principalmente pela estatística e econometria. [Atsalakis e Valavanis \(2009\)](#) apresentam um amplo estudo sobre estas técnicas;
- **Métodos do mercado:** estes podem ser subdivididos em dois tipos: a análise fundamentalista, sendo um dos grandes expoentes dessa escola o economista [Graham \(1917\)](#); e a análise técnica, cujos pais são os americanos [Hamilton e Dow \(1922\)](#), este último um dos fundadores da *Dow Jones* e do *Wall Street Journal*. *Charles Dow* formulou uma série de estudos que vieram a compor o núcleo da análise técnica, que ficou conhecida como a Teoria de *Dow*;
- **Sistemas computacionais:** em sua maioria fazem uso de inteligência computacional; [Tkáč e Verner \(2016\)](#), [Cavalcante et al. \(2016\)](#) apresentam uma extensa revisão da aplicação de diversos métodos computacionais em aplicações financeiras.

Nas seções seguintes abordaremos alguns trabalhos que fazem uso de inteligência compu-

tacional aplicada a séries temporais financeiras, meta-classificadores em geral e, por fim, trabalhos que utilizaram meta-classificadores para o mercado financeiro, respectivamente.

2.1.1 Inteligência Computacional em Séries Temporais Financeiras

Os autores [Nametala et al. \(2016\)](#) construíram uma estratégia automatizada (robô investidor) que combina previsões feitas por redes neurais artificiais e preditores econométricos em uma segunda rede neural; esta age como um comitê de decisão. Estas previsões foram utilizadas para gerar sinais de compra ou venda por meio de um modelo de negociação incorporado ao algoritmo. Os experimentos foram conduzidos com séries reais referentes a três ativos de alta liquidez, uma *commoditie* e um índice de mercado. Os resultados obtidos foram comparados de forma financeira frente à aplicação individual de cada preditor, e também, a técnicas clássicas de mercado. Os resultados mostraram que, para a maioria dos ativos, na média da carteira e nos retornos percentuais totais, os lucros obtidos pela combinação que o robô fez dos preditores geraram resultados melhores do que quando cada um deles foi aplicado individualmente.

[Persio e Honchar \(2016\)](#) aplicaram uma Rede Neural Artificial (ANN) ([WIDROW; HOFF, 1960](#)) para prever índices de mercado de ações. De acordo com os autores, foi elaborada uma abordagem inovadora baseada na combinação de *Wavelets*¹ ([STRUZIK, 2001](#)), Redes Neurais Convolutivas (CNN) ([LECUN et al., 1998](#)) e Redes Neurais Recorrentes (LSTM) ([HOCHREITER; SCHMIDHUBER, 1997](#)); essa abordagem superou as abordagens de redes neurais básicas. Testes foram realizados em séries financeiras históricas da *S&P500*².

Segundo [Giacomel, Galante e Pereira \(2015\)](#), embora muitos estudos de previsão de séries temporais estejam focados na previsão de valores exatos no futuro, evidências mostraram que esse tipo de problema funciona melhor quando o transformamos em um problema de classificação. Portanto, os autores propuseram um agente de negociação baseado em um *ensemble* de Redes Neurais que prevê se uma ação vai aumentar ou cair em vez de prever seus valores futuros. Para validar a eficiência do método proposto em diferentes situações, o algoritmo foi avaliado em dois conjuntos de dados diferentes: os mercados de ações norte-americano e brasileiro. Operações reais foram simuladas nesses mercados e a solução demonstrou capacidade de lucrar em todas as séries temporais testadas.

De acordo com [Kampouridis e Otero \(2015\)](#), a previsão financeira é uma área muito importante nas finanças computacionais. Eles desenvolveram o *EDDIE*, um algoritmo que utiliza a Programação Genética. A solução foi aplicada com sucesso em vários conjuntos de dados financeiros internacionais. Segundo os autores, o objetivo principal deste trabalho

¹ *Wavelet* ou ondaleta é uma função capaz de decompor e descrever uma série de dados originalmente descrita no domínio do tempo.

² O índice S&P 500 é composto por quinhentos ativos (ações) selecionados devido ao seu tamanho de mercado, sua liquidez e sua representação de grupo industrial.

foi melhorar ainda mais o desempenho preditivo do algoritmo, incorporando heurísticas na busca. Para examinar a eficácia de seu desempenho preditivo, a solução foi comparada com os algoritmos *C4.5*³ (QUINLAN, 1993) e *RIPPER*⁴ (COHEN, 1995). Os resultados mostram que a introdução de heurísticas foi muito bem sucedida, permitindo que o algoritmo superasse todos os algoritmos *C4.5* e *RIPPER*. Os resultados também mostraram que o algoritmo foi capaz de ter ganhos financeiros na maioria dos conjuntos de dados.

Cruz et al. (2015) desenvolveram um *ensemble* dinâmico denominado *META-DES* apresentando uma nova estrutura de meta-aprendizagem. Propuseram cinco conjuntos distintos de meta-características, cada uma correspondente a um critério diferente para medir o nível de qualidade de um classificador. As meta-características foram extraídas dos dados de treinamento e usadas para treinar um meta-classificador. Foram utilizados 30 conjuntos de dados extraídos do repositório *UCI* (LICHMAN, 2013). Os resultados apresentados foram promissores em sua maioria e obtiveram ganhos de acurácia quando comparados com *Bagging* e *AdaBoost*.

Patel et al. (2015) propuseram um algoritmo que faz a fusão de três algoritmos de aprendizado de máquina. A abordagem realizou a fusão em dois estágios, envolvendo a Regressão Vetorial de Suporte (*SVR*) (DRUCKER et al., 1997) na primeira etapa e, na segunda etapa, uma abordagem de Rede Neural Artificial (*ANN*), *Random Forest* (*RF*) (BREIMAN, 2001) e *SVR* resultando em modelos de predição de fusão *SVR-ANN*, *SVR-RF* e *SVR-SVR*. O algoritmo foi testado com dados reais da Bolsa de Valores de Bombaim (*BSE*), na Índia, utilizando 10 anos de dados históricos. O desempenho da predição destes modelos híbridos foi comparado com os cenários de um único estágio onde *ANN*, *RF* e *SVR* são usados sozinhos. A abordagem híbrida apresentou melhores resultados.

Em sua pesquisa, Dabhi e Chaudhary (2015) também apontaram que a previsão das séries temporais financeiras é considerada uma tarefa altamente desafiadora. Segundo eles, a tarefa torna-se difícil devido às características não lineares e não-estacionárias inerentes às séries temporais financeiras. Dessa forma, os autores propuseram uma combinação de *Wavelet* e *Postfix-GP*, um sistema de programação genética baseado na notação *postfix*⁵, para a previsão financeira de séries temporais. A capacidade de previsão foi testada em duas séries de preços de ações e dois índices de ações. Os resultados obtidos foram comparados com o *framework ECJ* (*Java-based Evolutionary Computation Research System*) (LUKE et al., 2007). Os resultados mostraram um desempenho melhor das previsões do *Postfix-GP*.

Em outro trabalho, Pimenta et al. (2014) apresentaram implementações de algoritmos Mono-Objetivo e Multi-Objetivo para a tomada de decisão e negociação de ações para a

³C4.5 é um algoritmo utilizado para criar uma árvore de decisão.

⁴RIPPER é um algoritmo de extração de regras.

⁵O termo *postfix* indica a posição dos operandos em relação ao operador; neste caso ficarão após. Exemplo da função: $X + Y$ ficaria $XY+$

bolsa de valores. Esses algoritmos utilizaram a Programação Genética (PG) combinada com os indicadores de análise técnica (médias móveis e índice de força relativa, entre outros), identificando o momento da compra e da venda. Os testes foram realizados com dados históricos da B3 com quatro ativos, BOVA11, PETR4, VALE5 e BBAS3. Os resultados mostraram-se promissores tanto na versão Mono quanto na Multi-Objetivo. Os autores ressaltaram que, mesmo com o mercado em queda, foi possível obter lucro.

Já [Seker et al. \(2013\)](#) aplicaram um *ensemble* em dados históricos financeiros a fim de encontrar correlação entre séries temporais financeiras e notícias da economia. Os algoritmos combinados foram *C4.5*, *SVM* e *KNN*. Os resultados mostraram ganhos de 3% na taxa de acerto quando comparados com os algoritmos individuais.

Outro problema identificado no segmento do mercado financeiro é a fraude. Para [Abbasi et al. \(2012\)](#), este tipo de problema pode ter consequências graves para a sustentabilidade em longo prazo de uma organização, bem como efeitos adversos sobre os seus empregados e investidores e sobre a economia como um todo. Segundo os autores, várias das maiores falências na história dos EUA foram de empresas que se envolveram em grandes fraudes. Por conseguinte, tem-se dado ênfase considerável ao desenvolvimento de abordagens automatizadas para a detecção de fraudes financeiras. Assim, os autores desenvolveram o *MetaFraud*, uma nova estrutura de meta-aprendizagem para a detecção de fraude financeira. Para validarem a estrutura proposta, uma série de experimentos foram conduzidos em um conjunto de dados de uma instituição financeira. Os resultados revelaram a eficácia do *framework* de meta-aprendizagem sobre os métodos atuais de detecção de fraude financeira do banco.

Os autores [Soltani, Akbari e Javan \(2012\)](#) apresentaram um modelo baseado em Sistemas Imunológicos Artificiais (SIA) ([FARMER; PACKARD; PERELSON, 1986](#)) para detecção de fraudes em séries financeiras. O modelo proposto analisa o comportamento da conta de cada cliente. Os resultados foram animadores e trouxeram boa taxa de acerto; no entanto, o algoritmo apresentou um longo tempo na fase de treinamento.

[Caldeira et al. \(2012\)](#) aplicaram técnicas de inteligência computacional para detectar fraudes em transações eletrônicas. As técnicas utilizadas foram *Random Forest*, Redes Neurais, *SVM* e Sistemas Imunológicos Artificiais. Dados históricos financeiros do UOL PagSeguro⁶ foram utilizados. A técnica que obteve melhores resultados foi a rede neural. Além disso, a técnica apresentou ganhos superiores ao praticado pela corporação.

⁶<https://pagseguro.uol.com.br/>

2.1.2 Meta-Classificadores em Geral

Talagala, Hyndman e Athanasopoulos (2018) propuseram uma estrutura para seleção de modelos de previsão usando meta-aprendizagem baseada em características de séries temporais. O algoritmo FFORMS (*Feature-based FORecast-model Selection*) proposto utilizou o conhecimento do desempenho passado de modelos candidatos de previsão em uma coleção de séries temporais para identificar o melhor método de previsão para uma nova série. Para validação, a solução foi aplicada em dados de competição M1 e M3 (MAKRIDAKIS; HIBON, 2000). Os autores demonstraram que o método apresentou melhores resultados que os métodos de *benchmark*.

Yerima e Sezer (2018) apresentaram um meta-classificador denominado *DroidFusion*. Segundo os autores, o número de *malwares*⁷ para *Android*⁸ vem crescendo consideravelmente. Por isso, eles desenvolveram uma abordagem que emprega a fusão de classificadores baseada em uma arquitetura multi-nível que permite uma combinação de algoritmos de aprendizado de máquina para obter uma maior precisão. O *framework* proposto possui uma camada de classificadores base em um primeiro nível e, em seguida, aplica um conjunto de algoritmos baseados em classificação em suas precisões preditivas no nível mais alto para obter um classificador final. Para validação de método, foram realizados experimentos em quatro conjuntos de dados de três coleções de amostras de aplicativos para *Android*. Os resultados mostraram-se promissores, visto que quando, comparado ao *Stacked Generalization (StackingC)* (SEEWALD, 2002), o *DroidFusion* apresentou melhor performance na maioria dos testes.

Conforme declararam os autores Mishra, Pokhrel e Hoque (2018), a identificação de proteínas de ligação a DNA apenas a partir de informação de sequência é um dos problemas mais desafiadores no campo da Bioinformática. A partir deste problema os autores desenvolveram uma técnica de meta-classificação baseada em *Stacking*, chamada *StackDPPred*, para a predição de proteínas de ligação a DNA, dada a sequência proteica. O objetivo da solução proposta foi melhorar a precisão da predição das proteínas de ligação ao DNA. Os autores concluíram que a técnica utilizada ajudou a reduzir o erro de generalização e, portanto, melhorou significativamente a precisão quando comparada com classificadores isolados.

2.1.3 Meta-Classificadores para Mercado Financeiro

Para os autores Mykova, Hajek e Olej (2018), existe um campo de grandes oportunidades na análise de texto para previsão de indicadores do mercado de ações, como preços de

⁷É um programa de computador destinado a infiltrar-se em um sistema de computador alheio de forma ilícita, com o intuito de causar alguns danos, alterações ou roubo de informações (confidenciais ou não)

⁸É o nome do sistema operacional baseado em Linux que opera em celulares (smartphones), netbooks e tablets.

ações, volumes de negociação e volatilidade. Desta forma, os mesmos desenvolveram um modelo de meta-aprendizagem para prever a volatilidade do retorno no mercado de ações que utiliza indicadores financeiros e artigos de notícias. A abordagem comparou sete meta-classificadores com três classificadores tradicionais com dados de quatorze ativos da Bolsa de Valores de Nova Iorque (NYSE). O resultado foi muito promissor, mostrando melhor performance dos meta-classificadores em grande parte dos conjuntos de dados.

Segundo [Duarte e Lorena \(2018\)](#), saber quando comprar (ou vender) um ativo é uma decisão importante para a obtenção de lucros em mercados financeiros. Entretanto, antes mesmo de se tomar tal decisão, um investidor deve escolher o modelo ideal para o estudo e análise de uma série de dados, devido à grande quantidade de modelos disponíveis. Desta forma, os autores acreditaram que a meta-aprendizagem seria uma ótima ferramenta para auxiliar em tal decisão. Eles buscaram utilizá-la para o estudo e análise de séries temporais, mais especificamente em dados do mercado Euro/Dólar, através de um algoritmo de ranqueamento que classifica e ordena os melhores modelos para o usuário. Como resultado, a aplicação de meta-aprendizagem mostrou-se muito promissora na sugestão de modelos para o estudo de séries temporais financeiras.

Como apresentado pelos autores [Barak, Arjmand e Ortobelli \(2017\)](#), prever os retornos das ações e seu risco representa uma das preocupações mais importantes dos tomadores de decisão de mercado. Muitos estudos encontrados na literatura têm utilizado classificadores únicos para a previsão de retornos de ativos e métodos de risco. Os métodos de fusão, que só surgiram recentemente, exigem um estudo mais aprofundado nesta área. Um conjunto de métodos, incluindo *Bagging* ([BREIMAN, 1996](#)), *Boosting* ([SCHAPIRE, 1990](#)) e *AdaBoost* ([FREUND, 1999](#)) foi aplicado para criar tal diversidade em combinações. Os resultados demonstraram que os meta-classificadores trouxeram desempenho superior através de testes realizados com dados da Bolsa de Valores de Teerã (TSE) do período de 2002 a 2012.

[Agapitos, Brabazon e O'Neill \(2017\)](#) desenvolveram um meta-classificador denominado *Regularised Gradient Boosting* (RGB), que é uma extensão do *Gradient Boosting* (GB). O objetivo desta proposta é diminuir o *overfitting* do classificador, resultando na eliminação de parâmetros de pouca importância. Para validação da performance da solução, o método foi aplicado ao domínio de séries temporais financeiras. Como resultado dos testes, a proposta superou o *Gradient Boosting* e mais dois métodos de aprendizado de máquinas na maioria dos conjuntos de dados utilizados.

De acordo com [Armaki et al. \(2017\)](#), praticamente todas as instituições financeiras estão expostas ao risco de crédito ao emitirem empréstimos ao consumidor. Desta forma, o desenvolvimento de soluções confiáveis de pontuação de crédito é um ponto crucial. Segundo os autores, diversas técnicas de aprendizado de máquina vêm demonstrando boa aplicabi-

lidade neste cenário. Porém, em diversos estudos, modelos híbridos através da fusão de vários algoritmos de aprendizado de máquina vêm se destacando frente aos algoritmos isolados. Diante deste cenário de oportunidades, os autores apresentaram um modelo de meta-aprendizado em que a solução utiliza três classificadores base (nível 0) para treinar o meta-classificador (nível 1). O classificador utilizado como meta foi o *Deep Belief Networks* (DBN), ou seja, uma rede neural profunda. Para validação da solução, foram utilizados quatro conjuntos de dados da *UCI Machine Learning Database Repository*. Os resultados foram muito positivos, pois em praticamente todos os cenários o meta-classificador obteve melhor performance que os classificadores isolados.

2.1.4 Considerações Finais

Diante dos trabalhos relacionados nesta pesquisa, os quais são direcionados para algoritmos de inteligência computacional, *ensembles*, fusões e meta-classificadores, percebe-se que ainda há muitas oportunidades para o mercado financeiro, possibilitando ainda mais espaço para explorar tal proposta e trazendo inovação e novas contribuições.

Sendo assim, esta tese de doutorado diferencia-se dos demais trabalhos relacionados aqui por realizar a implementação de uma solução que utiliza as predições dos classificadores base como atributos em um novo conjunto de treinamento. O processo funciona com uma arquitetura em camadas. Desta forma, cria-se um novo conjunto de treinamento, que é composto pelas predições dos classificadores do nível anterior. Esse novo conjunto de treinamento é utilizado para treinar o nível subsequente do sistema, também formado por classificadores, gerando assim meta-classificadores com o objetivo de fornecer a predição final de uma instância. Toda esta abordagem é encapsulada em um arcabouço metodológico que implementa uma modelagem e a avaliação de estratégias de operação, utilizando para isso conceitos de teorias econômicas, ciência de dados, análise estatística, análise de desempenho, análise de operação/risco, juntamente com os gráficos de evolução de capital acumulados.

Capítulo 3

Fundamentação Teórica

Neste capítulo serão detalhadas as principais abordagens que são de fundamental importância para o nosso tema de pesquisa, principalmente conceitos básicos em relação às séries temporais financeiras, ao mercado financeiro e a finanças computacionais. Também serão apresentados os algoritmos que foram utilizados no desenvolvimento do meta-classificador.

3.1 Séries Temporais Financeiras

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes. Algumas destas características são particulares a este tipo de dados, por exemplo as observações correlacionadas são mais difíceis de analisar e requerem técnicas específicas, precisando levar em conta a ordem temporal das observações. Fatores complicadores como a presença de tendências e variação sazonal ou cíclica podem ser difíceis de estimar ou remover. Além disso, a seleção de modelos pode ser bastante complicada e as ferramentas podem ser de difícil interpretação (BOX; JENKINS, 1976). A Figura 1 mostra um exemplo gráfico de uma Série Temporal Financeira.

Para Tsay (2005), a maioria dos estudos financeiros, ao invés de utilizarem os preços dos ativos, optam por utilizar os retornos gerados pelos mesmos, tendo razões fundamentais para justificar a utilização de retornos. As séries de retorno são mais fáceis de manusear do que a série de preços porque possuem propriedades estatísticas mais atraentes, como estacionariedade e ergodicidade.

Uma série temporal é dita estacionária quando a mesma se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Entende-se por ergodicidade a condição em que apenas uma realização do processo é suficiente para obter todas as estatísticas do mesmo.

Figura 1 – Exemplo de Série Temporal Financeira



Fonte: [Google](#) (2017)

Sendo assim, o retorno de um investimento ou ativo qualquer entre os períodos t e $t + 1$ é definido pela Equação (1):

$$r_t = \frac{p_t}{p_{t-1}} - 1, \quad (1)$$

onde p_t é o preço do investimento ou ativo no instante t .

Segundo [Tsay \(2005\)](#), também é muito comum em Finanças, quando temos dados diários, utilizar o chamado log-retorno, que é definido pela Equação (2):

$$l_t = \ln \left(\frac{p_t}{p_{t-1}} \right). \quad (2)$$

Os log-retornos possuem diversas vantagens com relação aos retornos: sob suposição de log-normalidade dos preços, o log-retorno possui distribuição normal, que é fácil de tratar; os log-retornos são aditivos ao longo do tempo, isto é, a soma dos log-retornos de períodos consecutivos é o log-retorno do período total. Em contraste, o retorno acumulado de vários períodos é o produtório de vários termos envolvendo cada retorno; O log-retorno é

aproximadamente igual ao retorno para períodos curtos. Assim, neste trabalho utilizaremos como principal *input* as séries de log-retorno.

Conhecendo o comportamento das séries financeiras e sabendo como tratá-las, na próxima seção será apresentado o funcionamento do mercado financeiro.

3.2 Mercado Financeiro

O mercado financeiro é o local onde podem ser negociados, comprados ou vendidos, bens como valores mobiliários, mercadorias e câmbio. A função do mercado é reunir diversos vendedores em um único lugar, tornando-os acessíveis aos compradores interessados. Mercados são considerados como parte vital de qualquer economia, pois, quanto mais ativos, mais se traduzem em oportunidades para que pessoas e empresas apliquem seus recursos e contribuam para o aquecimento da economia (ASSAF, 2009).

O princípio do mercado financeiro é simples, tentar comprar quando acredita-se que o ativo vai subir de preço no futuro e vender quando acredita-se que o preço irá diminuir, de forma a obter lucro. No entanto, realizar essas previsões é um desafio gigantesco, pois existem inúmeros fatores que podem influenciar o valor de um ativo.

O preço de uma ação em bolsa é fruto das condições de mercado (oferta e demanda) que refletem as condições estruturais e comportamentais da economia de um país, específicas da empresa e de seu setor econômico (FORTUNA, 2007).

Ações são títulos que representam frações de capital de uma determinada empresa. Uma ação é uma fração da empresa. Quando adquirimos ações de uma empresa, nos tornamos sócios desta, participando de seus resultados. Esses títulos são emitidos quando uma empresa toma a decisão de abrir seu capital para o público, e ser negociada na bolsa de valores (FILHO; ISHIKAWA, 2000). Existem dois tipos de ações (CAVALCANTE; MISUMI, 2005):

- **Ordinárias:** estas ações, além de proporcionarem a participação nos resultados da empresa, conferem o direito de voto nas assembleias da empresa.
- **Preferenciais:** adquirindo este tipo de ações, o investidor terá preferência no recebimento de dividendos, porém, não terá o direito a voto nas assembleias.

No Brasil, temos a Bolsa de Valores de São Paulo, que é responsável por gerir e administrar todas as negociações (BM&FBOVESPA, 2017). A todo momento existem agentes tentando comprar ou vender ativos e tendo suas decisões influenciadas pelos mais diversos fatores, desde análises bem fundamentadas até aspectos puramente psicológicos ou especulativos, causando grande volatilidade nos preços dos ativos (CUTLER; POTERBA; SUMMERS,

1988).

Em relação às previsões no mercado financeiro, dada a sua complexidade e dinamismo inatos, existe um constante debate quanto à possibilidade de prever mudanças de preços em ações. Fama e Malkiel (1970) introduziram a hipótese do “Mercado Eficiente” que define que o preço atual de um ativo sempre reflete toda a informação previamente existente instantaneamente, ou seja, não é possível utilizar nenhuma informação conhecida para prever preços futuros. Portanto, qualquer técnica de previsão de preços consiste em um esforço inútil.

Existe também a hipótese do “Passeio Aleatório” (MALKIEL, 1973) que alega que o preço de uma ação varia independentemente do seu histórico, em outras palavras, o preço de amanhã irá ser influenciado apenas pela informação de amanhã, independentemente do preço de hoje.

Essas duas hipóteses juntas determinam que não existem meios de prever precisamente o preço de uma ação e ainda existem experimentos que mostram que uma estratégia de investimento utilizando um agente tomando decisões completamente aleatórias pode por vezes superar alguns métodos de investimento como “Convergência e Divergência de Médias Móveis” (*Moving Average Convergence / Divergence*) e “Índice de Força Relativa” (*Relative Strength Index*) (BIONDO et al., 2013).

Por outro lado, existem diversos outros autores que alegam por meio de experimentos que, de fato, variações de preços de ações podem sim ser previstas até certo nível (LO; MACKINLAY, 1999). Vários métodos para previsão e modelagem do comportamento de ações são objeto de estudo de várias disciplinas, sendo relevante notar que em 2012 era estimado que aproximadamente 85% das negociações de ações feitas nas bolsas de valores dos Estados Unidos eram feitas através de algoritmos (GLANTZ; KISSELL, 2013).

Os métodos de análise do mercado financeiro podem se dividir em dois tipos (ANGHEL, 2013):

- **Análise fundamentalista:** acredita-se que o valor das ações de determinada companhia é determinado por fatores econômicos e pelo desempenho financeiro da própria companhia. Eventuais flutuações nos preços são corrigidas pelo mercado nos próximos períodos uma vez que a análise fundamentalista afirma que o preço de uma ação é sempre justo.
- **Análise técnica:** interpreta que toda a informação utilizável já está contida no valor das ações e, portanto, os únicos dados usados na análise do mercado podem ser os valores anteriores deste próprio mercado. A análise técnica também afirma que os movimentos do mercado que ocorrem no passado tendem a se repetir no futuro, o

que reforça a necessidade de observar estes valores passados quando se quer prever os futuros.

Este trabalho tem como foco principal a análise técnica. Desta forma, na próxima seção será abordada com maior propriedade e detalhamento.

3.2.1 Análise Técnica

Um dos métodos mais populares para a tomada de decisões de investimento na bolsa de valores é a análise técnica, que é definida por [Kirkpatrick e Dahlquist \(2006\)](#) como uma metodologia de previsão de preços baseada apenas em dados históricos do mercado, principalmente de preço e volume. A análise técnica também se baseia nas seguintes suposições:

- Fatos passados tendem a se repetir no futuro.
- Preços de ativos são definidos unicamente pela relação de oferta e demanda.
- Preços de ativos variam seguindo tendências.
- Alterações de oferta e demanda causam reversões nas tendências.
- Alterações de oferta e demanda podem ser identificadas em gráficos.
- Padrões em gráficos tendem a se repetir.
- O mercado possui memória.

De modo geral, investidores que utilizam esta análise técnica fazem uso de gráficos para detectar possíveis tendências. Acredita-se que estas são baseadas na oferta e procura, causando de modo geral um padrão cíclico ([NORONHA, 2003](#)). Existem inúmeros indicadores técnicos que permitem este tipo de análise. Estes podem ser utilizados também como entrada para os algoritmos de inteligência computacional ou até para modelos de estratégias de negociação ([PIMENTA et al., 2014](#)). Abaixo estão listados alguns tipos de indicadores juntamente com seus exemplos ([INVESTIDOR, 2017](#)).

- **Indicadores de Momento:** normalmente fornecem pontos (momentos) de entrada e saída (ex.: Índice de Força Relativa, *Williams %R*, *ROC*);
- **Indicadores de Tendência:** fornecem a direção do mercado; se está subindo ou descendo (ex.: Média Móvel, *ADX*, *MACD*);
- **Indicadores de Volatilidade:** mostram se os preços estão muito voláteis; subindo e descendo sem tendência definida (ex.: *Bollinger Band*, Canal de Preço, *Average True Range*);
- **Indicadores de Volume:** são baseados no fato de que o volume precede o movimento do preço (ex.: *OBV*, Oscilador de Volume, *Chaikin Oscilador*).

Normalmente para criar os indicadores são necessárias informações contidas em um *can-*

dlestick (GRAF BOLSA, 2017). Essas informações são: os preços de abertura, fechamento, máximo e mínimo.

- **Fechamento:** mostra o valor que a ação atingiu ao final do período. É a métrica mais conhecida e utilizada na análise do mercado;
- **Abertura:** é o preço da ação no momento da abertura do pregão. Não é necessariamente igual ao valor de fechamento do dia anterior, visto que geralmente uma ação começa o dia já em alta ou em queda, devido a operações de *pre-market*, feitas antes do horário de abertura do pregão, e de *after-market*, feitas após o final do pregão anterior;
- **Máximo:** geralmente, ao longo do dia, o preço de uma ação sofre variações até chegar no valor de fechamento. Esta métrica mostra o valor máximo atingido pela ação ao longo do período. Este valor ajuda na interpretação de um preço porque quando o valor máximo de uma ação é muito maior do que o valor de fechamento, significa que o preço da ação caiu ao longo do período e a tendência é de que continue caindo no período seguinte. O inverso pode ser interpretado para um valor máximo muito perto do valor de fechamento;
- **Mínimo:** mostra o menor valor atingido por uma ação durante o dia. Sua interpretação é inversa à do valor máximo;
- **Volume:** é a soma do valor envolvido em todas as operações de compra e venda de determinada ação ao longo do dia.

Um *candlestick* é uma forma retangular verticalizada e padronizada que identifica os preços de uma ação, ou qualquer ativo financeiro negociado no mercado. Nas extremidades do retângulo pode haver linhas que saem da caixa, perpendicularmente ao retângulo.

As extremidades destas linhas representam o preço máximo ofertado pelo ativo, se for acima do retângulo, e o preço mínimo, se for abaixo. Estas linhas também são chamadas de sombra. Por isso, convencionou-se identificá-lo como *Candle Stick*, que, traduzindo significa bastão de vela, por realmente se parecer com uma vela. Abaixo, é possível demonstrar os dois tipos básicos de *candle*, o de queda, ou seja, variação negativa dos preços no período de tempo, e de subida, de variação positiva. As cores podem ser editadas e dependem das preferências de cada um, mas, geralmente, utiliza-se vermelho ou preto para *candles* de baixa, e verde ou branco para *candles* de alta.

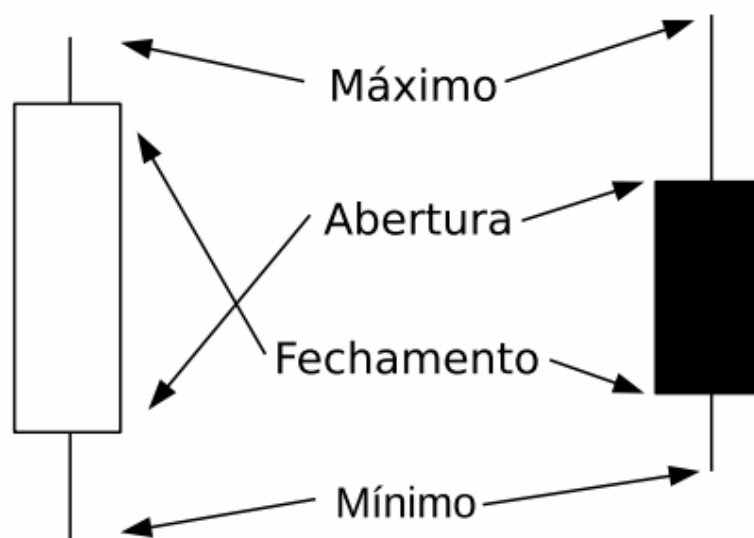
O corpo do *candle* representa a diferença entre o preço de abertura e o preço de fechamento e, dependendo do mercado, pode se configurar de vários modos: desde corpos muito longos até corpos estreitos nos quais não se pode identificar a diferença entre os preços. Neste caso, o *candle* é denominado *Doji*¹. Conforme os preços variam no mercado, os *candles*

¹Doji é o nome de um padrão de análise gráfica de *Candlestick*; uma técnica usada para análise de mercado.

vão se dispendo lado a lado, formando variadas configurações ao longo de um período de tempo.

Algumas disposições podem ser interpretadas como potenciais alertas para reversões e continuações de tendências, para uma possível disparada nos preços e para um bom sinal de compra ou venda (INVESTIDOR, 2017). A Figura 3 mostra um gráfico de *candles* de um determinado ativo.

Figura 2 – *Candlestick*.



Fonte: Autor

3.2.2 Estratégias de Negociação

Segundo Damien (2018), com o desenvolvimento das plataformas de negociação *on-line*, tornou-se possível para qualquer pessoa especular diretamente nos mercados financeiros e tomar posições para a compra ou venda dos mais diversos ativos. Entretanto, existem duas maneiras de negociar *on-line*. A primeira consiste em operar manualmente, desenvolvendo estratégias próprias, e a segunda que é chamada “*trading* por algoritmos” e consiste em um modo de operação automatizado.

O *trading* por algoritmos, também chamado *algotrading*, utiliza um sistema de cálculos baseados em modelagem matemática. Trata-se, portanto, de um *software* que determina as posições a serem tomadas, normalmente utilizando o sentido da tendência, o volume a ser negociado e, finalmente, a estratégia a ser adotada (DAMIEN, 2018). O *algotrading* apresenta certas características:

Primeiramente, permite aos investidores que não dispõem de tempo suficiente para dedicar à especulação nos mercados tomarem posições sem precisar acompanhar o mercado. Em

Figura 3 – Gráfico de *Candlesticks*.



Fonte: [Investidor \(2017\)](#)

segundo lugar, permite uma presença nos mercados 7 dias por semana e 24 horas por dia, o que nenhum ser humano poderia fazer. Finalmente, o *algotrading* segue uma estratégia rigorosa, exclusivamente fundamentada na análise técnica dos mercados, e não sofre de nenhum estresse ou emoção passível de favorecer erros de avaliação.

O *algotrading* pode ser utilizado só ou em complemento a uma análise humana da situação de um mercado. Utilizado sozinho, ele se constitui como um software de *trading* automático, posicionando as ordens no lugar do humano, de acordo com os sinais detectados. Em contrapartida, utilizado em complemento a uma análise fundamentalista, ele constitui apenas uma ajuda à análise global e para a tomada de decisão, favorecendo uma estratégia específica.

O investimento direto em ações e outros ativos de renda variável envolve alguns custos operacionais, tais como taxa de corretagem, taxa de custódia, emolumentos e imposto de renda que podem sofrer variações de valores conforme o tipo de transação realizada ([VIEIRA, 2015](#)). Esses custos serão detalhados a seguir.

- **Taxa de corretagem:** é um valor cobrado pelas corretoras para intermediar as operações de compra ou venda de ações junto à bolsa de valores. Esta taxa pode ser um valor fixo independente do volume da operação ou uma porcentagem sobre o montante da operação, variando de acordo com o volume transacionado acrescida do

Imposto Sobre Serviços (ISS) sobre o valor da corretagem. Esta taxa varia de acordo com cada município e costuma ser de no máximo 5%.

- **Emolumentos e liquidação:** os emolumentos são cobrados pela bolsa de valores (Bovespa) através de uma taxa fixa para cada tipo de operação, com a finalidade de cobrir custos operacionais que incidem na compra e venda direta de ações. As operações normais, compra e venda de ações efetuadas em pregões distintos, tem incidência de 0,0325%, sendo 0,0275% de Liquidação e 0,005% de Emolumentos. Já as operações de daytrade, compra e venda de um ativo em um mesmo dia, têm a incidência de 0,025%, sendo 0,02% de Liquidação e 0,005% de Emolumentos.
- **Taxa de custódia:** é um valor cobrado pelas corretoras para intermediar as operações de compra ou venda de ações junto à bolsa de valores. Dependendo da corretora ou do pacote de corretagem oferecido pela instituição financeira, o cliente pode até ficar isento da taxa de custódia.
- **Imposto de Renda:** o lucro obtido ao se operar renda variável, o chamado ganho líquido, está sujeito à cobrança de imposto de renda. A definição de ganho líquido é “a diferença positiva entre o valor de venda e o valor de compra de uma ação, já descontadas as taxas”. Ganhos líquidos com operações comuns são tributados em 15%. Já aqueles obtidos com operações *daytrade* (compra e venda de um ativo no mesmo dia) são tributados em 20%. Há isenção de IR para vendas no valor de até 20 mil reais em um único mês. Dividendos também são isentos de IR, assim como os rendimentos pagos por fundos imobiliários. Juros sobre Capital Próprio são tributados na fonte em 15%.

A apuração e o recolhimento do imposto de renda sobre o ganho líquido na negociação de ativos de renda variável são de responsabilidade do próprio investidor. Quando houver uma venda sujeita à tributação, o investidor deve emitir um Documento de Arrecadação de Receitas Federais (DARF) e pagá-lo até o último dia útil do mês seguinte ao da operação. Todas as tarifas cobradas, para todos os mercados e tipos de ativo, estão disponíveis no site da B3.

3.3 Inteligência Computacional, Ferramentas e Algoritmos

A Inteligência Computacional (IC), denominada originalmente de Inteligência Artificial (IA), é uma das ciências mais recentes, tendo surgido logo após a Segunda Guerra Mundial e tendo seu nome original cunhado em 1956 (GOLDSCHMIDT, 2010). A Inteligência Computacional vai além da perspectiva de compreensão do pensamento humano, pois também procura construir entidades artificiais inteligentes. IC é uma área da Ciência da Computação que tem como objetivos desenvolver e aplicar recursos que permitam aos computadores terem um comportamento similar ao humano em tarefas específicas. As atividades na área de IC

podem ser organizadas em três grandes grupos: atividades voltadas ao desenvolvimento tecnológico em IC, atividades de construção de sistemas inteligentes com recursos de IC e atividades envolvendo a aplicação de sistemas inteligentes (GOLDSCHMIDT, 2010).

Nas Seções seguintes detalharemos a ferramenta que compõe os algoritmos de IC utilizados no meta-classificador, além de detalhar cada um deles.

3.3.1 *Waikato Environment for Knowledge Analysis (WEKA)*

Waikato Environment for Knowledge Analysis (WEKA) (WITTEN et al., 1999) é uma suíte de mineração de dados muito popular no meio acadêmico, criada nas dependências da Universidade de *Waikato*, Nova Zelândia. Atualmente, é mantida por uma comunidade de entusiastas por ser um *software* livre disponível sobre a licença GPL.

O *Weka* está implementado na linguagem Java, que tem como principal característica a sua portabilidade. Desta forma, é possível utilizá-la em diferentes sistemas operacionais, além de aproveitar os principais benefícios da orientação a objetos. A ferramenta também possui quatro tipos métodos de aprendizado: agrupamento, associação, predição numérica e classificação. Este trabalho baseia-se no método de classificação.

3.3.2 Classificação

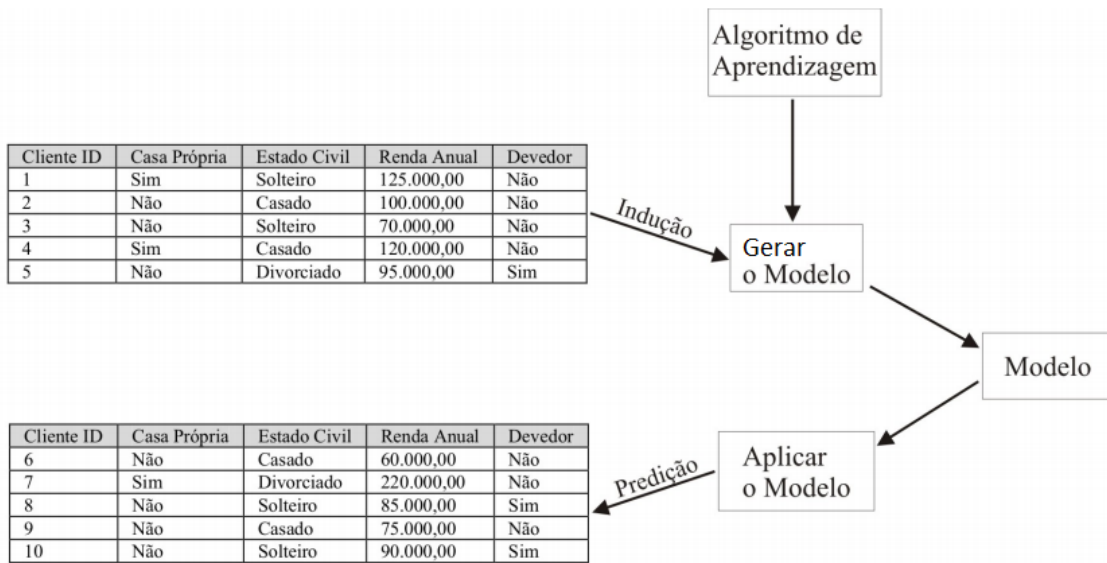
Para Fayyad, Piatetsky-Shapiro e Smyth (1996), classificação é o processo de encontrar um modelo (ou conjunto de funções) que descreve e distingue classes ou conceitos de dados. De posse deste modelo, é possível identificar objetos cuja classe ainda não é conhecida. A Figura 4 ilustra um processo genérico de classificação.

3.3.3 Máquinas de Vetores de Suporte (SVM)

As Máquinas de Vetores de Suporte (SVM) são embasadas pela teoria de aprendizado estatístico, desenvolvida por Vapnik (1995) a partir de estudos iniciados em Vapnik e Chervonenkis (1971). Esse estudo estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa generalização, que é definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu. Os algoritmos de aprendizagem de máquina SVM têm como objetivo a determinação de limites de decisão que produzam uma separação ótima entre classes por meio da minimização dos erros. As SVM destacam-se por ao menos duas características: sólida fundamentação teórica e alto desempenho em aplicações práticas (SANTOS, 2002).

Em sua forma básica, SVMs são classificadores lineares que separam os dados em duas classes através de um hiperplano de separação. Um hiperplano ótimo separa os dados com a máxima margem possível, que é definida pela soma das distâncias entre os pontos

Figura 4 – Processo de classificação.

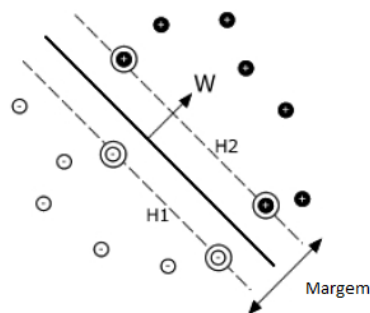


Fonte: [Basgalupp \(2007\)](#)

positivos e os pontos negativos mais próximos do hiperplano. Estes pontos são chamados de vetores de suporte e estão circutados na Figura 5.

O hiperplano é construído com base em treinamento prévio em um conjunto finito de dados.

Figura 5 – Classificação de um conjunto de dados usando uma SVM linear.



Fonte: [Readhead \(2017\)](#)

Assumindo o conjunto de treinamento $\{x_i, y_i\}$, $y_i \in \{-1, 1\}$, $x_i \in \mathbb{R}^n$, onde x_i é o i -ésimo elemento de entrada e y_i é o seu respectivo valor de classe para x_i , $i = 1, \dots, l$. O cálculo do hiperplano com margem ótima é dado pela minimização de $\|w\|^2$ obedecendo às seguintes restrições:

$$x_i w + b \geq +1, \quad y_i = +1 \quad (3)$$

$$x_i w + b \leq -1, \quad y_i = -1 \quad (4)$$

onde w é o vetor normal ao hiperplano, $\|w\|$ é uma norma de w e $|b|/\|w\|$ é a distância perpendicular do hiperplano da origem. Este é um problema quadrático de otimização e pode ser transformado para a forma dual, em que depende apenas dos Multiplicadores de Lagrange α_i :

$$u = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \mathbf{x}_j) \quad (5)$$

respeitando as restrições da equação linear:

$$\sum_{i=1}^N \alpha_i y_i = 0, \quad (6)$$

e as restrições da inequação:

$$\alpha_i \geq 0, \quad \forall i \quad (7)$$

com a solução dada por:

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (8)$$

Onde N é o número de exemplos de treinamento. Os elementos mais próximos ao hiperplano são chamados de vetores de suporte e estão localizados nos planos $H1$ e $H2$, como apresenta a Figura 5. Esses são os pontos mais importantes, pois são eles que definem a margem de classificação da SVM (BURGES, 1998).

Para a maioria dos problemas reais, o conjunto de dados não é separável através de um hiperplano linear, e o cálculo dos vetores de suporte utilizando as formulações descritas acima não se aplicaria (PLATT, 1999).

Tal problema pode ser resolvido através da função kernel. Alguns dos kernels mais utilizados na prática são os Polinomiais, os Gaussianos ou RBF (*Radial-Basis Function*) e os Sigmoidais.

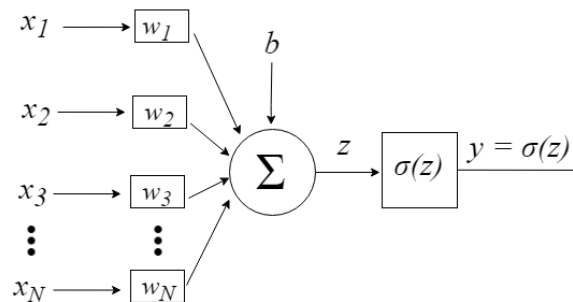
A SVM possui alguns hiper-parâmetros a serem escolhidos: a função kernel, gamma e o cost. As funções de kernel são responsáveis por proverem uma ponte simples entre algoritmos lineares e não lineares. O parâmetro cost determina o equilíbrio entre erros de treinamento e margens de separação, com o intuito de permitir flexibilidade na separação das classes. Já o parâmetro gamma do kernel RBF controla a flexibilidade do classificador (BEN-HUR; WESTON, 2010).

3.3.4 Redes Neurais Artificiais (RNA)

A história das Redes Neurais Artificiais (RNA) é bem extensa e há uma imensidão de pesquisadores envolvidos no assunto (WIDROW; HOFF, 1960; MCCULLOCH; PITTS, 1943; HEBB, 1949; ASHBY, 1960; WIDROW; HOFF, 1962; ROSENBLATT, 1958; WERBOS, 1994).

Em resumo, RNAs são modelos matemáticos que se assemelham às estruturas neurais biológicas e que têm capacidade computacional adquirida por meio de aprendizagem e generalização. Esses modelos almejam semelhança com o sistema nervoso dos seres vivos e a com sua capacidade de processar informações (ROSENBLATT, 1958). Estabeleceu-se na área da Inteligência Artificial um modelo computacional de um neurônio, conforme ilustrado na Figura 6.

Figura 6 – Modelo computacional de um Neurônio Artificial.



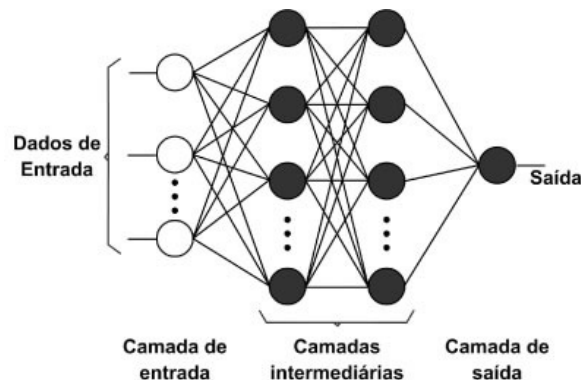
Os sinais da entrada no neurônio são representados pelo vetor $x = [x_1, x_2, x_3, \dots, x_N]$, podendo corresponder aos preços de um ativo, por exemplo. Ao chegarem ao neurônio, são multiplicados pelos respectivos pesos sinápticos, que são os elementos do vetor $w = [w_1, w_2, w_3, \dots, w_N]$, gerando o valor z , comumente denominado potencial de ativação, de acordo com a Equação 9.

$$z = \sum_{i=1}^N x_i w_i + b \quad (9)$$

O termo adicional b provê um grau de liberdade a mais, que não é afetado pela entrada nessa expressão, correspondendo tipicamente ao *bias* (viés). O valor z passa então por uma função matemática de ativação σ , com a característica de ser não linear, responsável por limitar tal valor a um certo intervalo, produzindo o valor final de saída y do neurônio. Algumas funções de ativação usadas são a degrau, a sigmoide, a tangente hiperbólica, a *softmax* e a *ReLU* (*Rectified Linear Unit*).

Segundo Ramos (2003), o algoritmo *Multilayer Perceptron* (MLP), que baseia-se na aprendizagem por correção de erros. Este algoritmo utiliza um tipo de aprendizado supervisionado. Quando o valor de saída é gerado, o erro é calculado e seus valores são retro-propagados para entrada, os pesos são ajustados e os valores são novamente calculados. A Figura 7 mostra um exemplo de MLP.

Figura 7 – Exemplo de uma Rede Neural Artificial Multilayer Perceptron (MLP).



Fonte: Pedregosa et al. (2011)

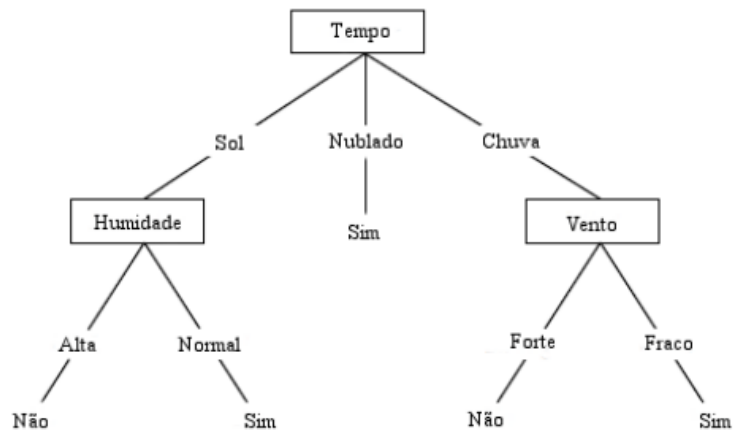
3.3.5 Árvore de Decisão (J48)

Uma Árvore de Decisão é um modelo de classificação/regressão cuja estrutura consiste em um determinado número de nós e arcos (também conhecidos como ramos) (QUINLAN, 1993).

O algoritmo J48 (C4.5) (QUINLAN, 1993) é um dos mais clássicos algoritmos de árvore de decisão. Este algoritmo utiliza o método divisão e conquista para aumentar a capacidade de predição das árvores de decisão. Assim, sempre usa o melhor passo avaliado localmente, sem se preocupar se esse passo vai produzir a melhor solução. Um problema é dividido em vários subproblemas, sendo criadas sub-árvores entre a raiz e as folhas. A Figura 8 mostra um exemplo de representação de uma J48.

A principal tarefa realizada pelo algoritmo é verificar qual característica que será utilizada como teste em cada nó. Naturalmente, o interessante é escolher uma característica que

Figura 8 – Exemplo de uma saída de Árvore de Decisão (J48).



Fonte: Adaptado de [Mitchell \(1997\)](#)

melhor classifique os subconjuntos que serão criados ao realizar o teste no nó. Este critério de seleção é calculado, e a característica com melhor resultado é utilizada como teste no nó.

As árvores de decisão classificam instâncias partindo da raiz da árvore para algum nodo folha que fornece a classe da instância. Cada nodo da árvore especifica o teste de algum atributo da instância, e cada arco alternativo que desce daquele nodo corresponde a um dos possíveis valores deste atributo. Uma instância é classificada começando no nodo raiz da árvore, testa o atributo relacionado a este nodo, e segue o arco que corresponde ao valor do atributo na instância em questão. Este processo é repetido, então, para a sub-árvore abaixo até chegar a um nodo folha.

3.3.6 Otimização Mínima Sequencial (SMO)

O algoritmo de Otimização Mínima Sequencial (SMO) é um SVM e foi proposto por [Platt \(1998\)](#). O SMO se propõe a resolver o problema de Programação Quadrática (PQ) que existe no treinamento do SVM (ver seção [3.3.3](#)).

A solução proposta pelo SMO é a decomposição do problema de PQ global em uma série de subproblemas e, então, escolher o menor problema de otimização possível para resolver em cada etapa. Estes pequenos sub-problemas são então solucionados de forma analítica, evitando assim o uso de métodos numéricos de otimização, computacionalmente mais intensivos ([PLATT, 1998](#)). O SMO lida particularmente bem com conjuntos de dados esparsos, bem como com dados de entrada binários ou não-binários.

3.3.7 Florestas Aleatórias (RF)

Florestas Aleatórias (*Random Forest* - RF) é uma técnica de aprendizado de máquina desenvolvida por Breiman (2001). Uma RF é uma combinação de árvores de decisão em que cada árvore depende dos valores de vetores aleatórios amostrados de forma independente e distribuídos igualmente para todas as árvores na floresta. Nesse método, depois que um determinado número de árvores é gerado, cada uma lança um voto para uma classe do problema, considerando um vetor de entrada. Então, a classe mais votada será escolhida na predição do classificador (BREIMAN, 2001). Essa técnica permite que se obtenha modelos muito eficazes sem nenhuma preparação de dados ou conhecimento de modelagem (GISLASON; BENEDIKTSSON; SVEINSSON, 2006).

3.3.8 Redes Bayesianas (BN)

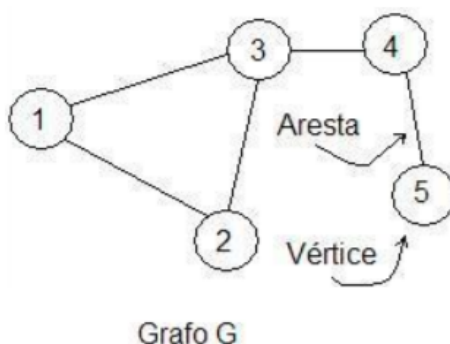
Redes Bayesianas (RB) são modelos de representação do conhecimento que trabalham com o conhecimento incerto e incompleto por meio do Teorema de Bayes, publicado pelo matemático Bayes (1763) e foram desenvolvidas no início dos anos 1980 por Pearl (1988).

As RB são grafos acíclicos e direcionados, sendo uma representação gráfica de variáveis e suas relações, dado um contexto específico. A RB é um dos principais tipos de modelos gráficos probabilísticos. Cada vértice representa uma variável aleatória e as arestas representam as correlações entre as variáveis.

Formalmente, um grafo G (Figura 9) é um par de conjuntos (V, E) em que:

- V é um conjunto não vazio cujos elementos são chamados vértices ou nós;
- E é um conjunto de pares dos vértices de V , cujos elementos são denominados arestas.

Figura 9 – Grafo G com ordem e tamanho igual a 5.



Para a correta compreensão de uma RB, é importante entender o que são grafos direcionados e não direcionados. Se as arestas são constituídas de pares ordenados de vértices,

diz-se que o grafo é direcionado.

Em um grafo direcionado, se existe uma aresta de v_i para v_j , diz-se que v_i é pai de v_j . Se existe um caminho direcionado de v_i para v_j , diz-se que v_i é ancestral de v_j . Se um grafo direcionado não possuir ciclos direcionados, isto é, se para qualquer vértice $v_i \in V$ não existe um caminho direcionado que começa e termina em v_i , então diz-se que é um grafo acíclico direcionado - DAG (*Directed Acyclic Graph*).

Os grafos direcionados acíclicos em RB são chamados de estrutura, e os valores da distribuição de probabilidades condicionais são chamados de parâmetros. Tanto a estrutura quanto os parâmetros de uma RB podem ser definidos manualmente, com os relacionamentos entre as variáveis sendo definidos tanto por especialistas, a partir de uma base de dados, quanto pela combinação de ambas as abordagens.

Segundo [Neapolitan \(2003\)](#), as RB são estruturas gráficas para representar as relações probabilísticas entre um grande número de variáveis e para fazer inferência estatística com essas variáveis. As RB permitem eficiente e efetiva representação da distribuição da probabilidade conjunta sobre um grupo de variáveis aleatórias. O objetivo dos modelos gráficos probabilísticos é criar uma estrutura matemática que une grafos e probabilidades e que permita modelar situações complexas envolvendo aleatoriedade ou incerteza.

A regra da cadeia vale tanto para distribuições condicionais discretas quanto para contínuas se a condição de *Markov* é satisfeita. RBs com variáveis discretas satisfazem a condição de *Markov* ([NEAPOLITAN, 2003](#)), que é dada por: cada variável da RB é condicionalmente independente do conjunto de todos os não-descendentes desta, dado o conjunto de todos os seus pais.

Um importante aspecto de uma RB refere-se à sua estrutura (topologia do grafo), a qual permite a representação de complexas relações entre variáveis de forma gráfica e intuitiva. A estrutura gráfica de uma RB facilita o entendimento das relações entre variáveis do seu domínio, além de permitir o uso combinado de informações obtidas do conhecimento de especialistas e de dados históricos para obter a distribuição conjunta de probabilidades da rede.

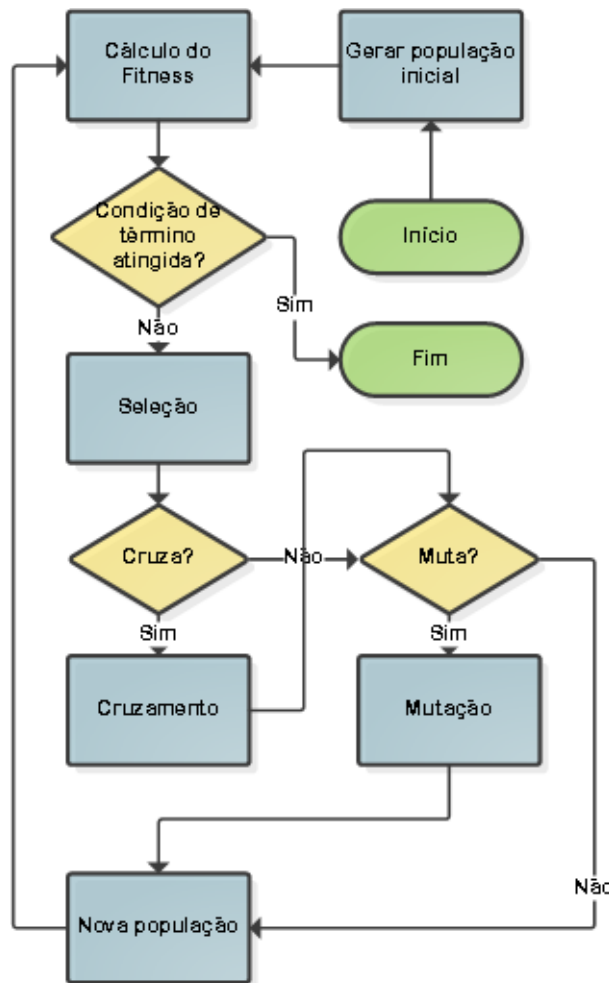
3.3.9 Programação Genética (PG)

A Programação Genética (PG) é um método de Inteligência Computacional que aplica os conceitos da biologia genética, baseados na teoria da seleção natural. O paradigma da PG foi desenvolvido por [Koza \(1992\)](#) com base na técnica de Algoritmos Genéticos (AG), desenvolvida por [Holland \(1975\)](#), utilizada por [Goldberg \(1989\)](#) e [Banzhaf et al. \(1998\)](#) em diversas aplicações de PG.

Segundo [Koza \(1992\)](#), a diferença principal entre o AG e a PG está na representação dos indivíduos: enquanto o AG codifica os indivíduos em código binário ou real, a PG codifica os indivíduos na forma de uma árvore. Essa codificação dos indivíduos da PG gera alterações significativas nos operadores genéticos, além de possibilitar grandes benefícios, tais como conter não somente valores de variáveis, mas também funções.

O algoritmo básico da PG é apresentado na Figura 10. O primeiro procedimento cria uma população de indivíduos. Na etapa de avaliação (função *fitness*), cada indivíduo é executado e um valor de qualidade do indivíduo é determinado. Caso o critério de parada seja satisfeito, o processo de evolução é finalizado. Caso contrário, indivíduos de melhor qualidade são selecionados com maior probabilidade para serem submetidos aos operadores genéticos, gerando novas soluções. Estas novas soluções irão compor uma nova população de indivíduos, a qual será submetida aos mesmos procedimentos, com exceção do primeiro que cria a população inicial. A execução iterativa desses procedimentos permite a emergência de boas soluções a cada geração.

Figura 10 – Fluxograma do funcionamento básico de um Programa Genético.



Fonte: Adaptado de [Koza \(1992\)](#)

O detalhamento de todo este fluxo será descrito nas subseções 3.3.9.1 e 3.3.9.2.

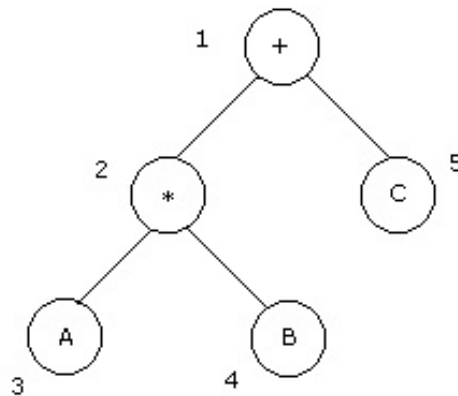
3.3.9.1 Indivíduo

O indivíduo é um ponto no espaço de otimização, isto é, um conjunto de valores das variáveis candidatas a otimizar a função objetivo (KOZA, 1992). O indivíduo representa uma solução candidata para o problema e sua codificação tem formato de árvore, como já mencionado.

Para Koza (1992), o fato do indivíduo na PG ser representado na forma de árvore possibilita a esse indivíduo conter não somente valores de variáveis, mas também funções. Esse conjunto que compõe o indivíduo é chamado de conjunto de funções.

Montana (1994) considera que o termo que melhor se aplica a esse conjunto é conjunto não terminal, uma vez que o conjunto de nós terminais pode conter funções que não recebem parâmetros. Esse trabalho utilizará o termo conjunto de funções, por ser o mais utilizado na literatura, devido à grande influência exercida pelo livro de Koza (1992). A Figura 11 ilustra um exemplo de indivíduo da PG.

Figura 11 – Exemplo de indivíduo na Programação Genética.



Fonte: Adaptado de Koza (1992)

Nesta representação, operadores e operandos estão dispostos em nós de uma árvore. A execução de um dado programa pode ser realizada de duas formas: percorrendo os nós em pós-ordem ou em pré-ordem. No primeiro caso, o programa é executado percorrendo os nós das extremidades até o nó raiz, ou seja, visitando primeiramente os operandos e em seguida os operadores (KOZA, 1992). No segundo caso, é realizado o processo inverso. Parte-se da raiz para cada ramo, os nós são visitados até chegar aos nós folha. Esta última abordagem tem a vantagem de economizar recursos computacionais na existência de nós condicionais, uma vez que só serão executados os nós da subárvore associada à condição satisfeita (KOZA, 1992).

O indivíduo da PG deve satisfazer pelo menos duas propriedades (KOZA, 1992):

- **Suficiência:** garante a convergência do sistema, fazendo com que os conjuntos de nós funções e nós terminais sejam capazes de representar uma solução viável para o problema em questão.
- **Fechamento:** garante que qualquer função do conjunto de nós funções deve ser capaz de operar com todos os valores recebidos como entrada. Isso garante que sejam geradas árvores sintaticamente viáveis.

Na prática, obedecer à propriedade de fechamento pode ser uma tarefa muito difícil. Entretanto, não observar essa propriedade leva a situações que geram indivíduos inconsistentes, ou seja, não faz sentido existir operações entre atributos de domínios diferentes. Certamente essa é uma das maiores dificuldades dos projetistas de algoritmos baseados em PG, principalmente na aplicação das operações de mutação e cruzamento (PEREIRA, 2012).

O conjunto de indivíduos é denominado população e a geração inicial, na maioria das vezes, é feita de forma aleatória (KOZA, 1992). A quantidade de indivíduos da população é um parâmetro de entrada da PG.

3.3.9.2 Operadores Genéticos

O princípio básico dos operadores genéticos é transformar a população através de sucessivas gerações, estendendo a busca até chegar a um resultado satisfatório (KOZA, 1992). Os operadores genéticos são necessários para que a população se diversifique e mantenha características de adaptação adquiridas pelas gerações anteriores.

Nas próximas seções serão descritos os três operadores genéticos utilizados: seleção, cruzamento e mutação, os quais são usados para modificar estruturas sob adaptação em PG.

3.3.9.3 Seleção

No processo evolutivo, a seleção de indivíduos é realizada pelo conceito de aptidão. Em PG, a aptidão é avaliada a partir dos valores obtidos da função *fitness*. Esta função consiste na mensuração da aptidão de cada indivíduo a partir da função objetivo que representa o problema (GOLDBERG, 1989). No problema do caixeiro viajante, por exemplo, uma função objetivo simples poderia ser o inverso da distância total percorrida. Desta forma, as rotas mais curtas teriam uma aptidão maior.

No processo de seleção, a técnica de *roulette wheel* realiza a escolha dos indivíduos proporcionalmente à sua respectiva aptidão (GOLDBERG, 1989), a Equação 10 modela a probabilidade de cada indivíduo ser selecionado. Onde p_i é a probabilidade do indivíduo

i ser escolhido pela roleta e f_i é a *fitness* do indivíduo. Esse procedimento de seleção é realizado uma única vez para cada indivíduo, o que representa o problema da seleção privilegiar indivíduos que possuam maior aptidão.

$$p_i = \frac{f_i}{\sum f_i} \quad (10)$$

Adicionalmente, a seleção pode ser implementada por torneio e elitismo. No torneio, são selecionados diversos pequenos subconjuntos da população, sendo selecionado o indivíduo de maior aptidão de cada um desses grupos. Maiores detalhes sobre *roulette wheel* e torneio podem ser vistos em [Goldberg \(1989\)](#).

Já a técnica de elitismo, consiste em manter o indivíduo mais apto para as gerações seguintes. O indivíduo mais apto é aquele que possui maior valor da função *fitness* ([JONG, 1975](#)).

3.3.9.4 Cruzamento

A operação de cruzamento na PG consiste na troca de material genético entre os indivíduos pais. Uma característica bastante significativa da operação de cruzamento é a sua capacidade de alterar os tamanhos dos indivíduos durante a execução do algoritmo ([PEREIRA, 2012](#)).

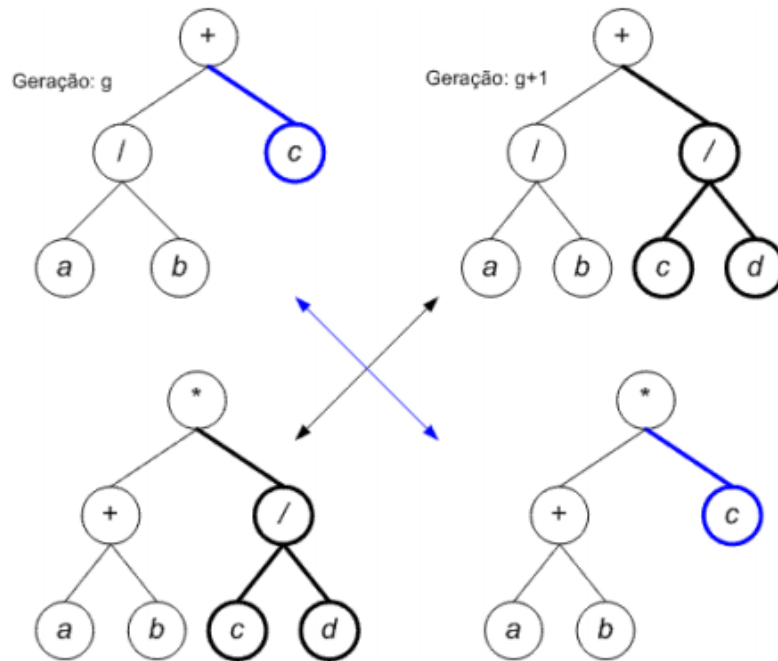
A operação de cruzamento, em relação ao conceito biológico, corresponde ao processo de reprodução sexuada ([GOLDBERG, 1989](#)). Dois indivíduos são selecionados e recombinados para gerar outros dois indivíduos a partir de seus valores de *fitness* ([KOZA, 1992](#)). Um ponto aleatório de cruzamento é escolhido em cada indivíduo-pai e as árvores abaixo destes pontos são trocadas. A probabilidade dos indivíduos-pais cruzarem é um parâmetro de entrada da PG.

Um exemplo de cruzamento pode ser visualizado na Figura 12.

3.3.9.5 Mutação

Como apresentado por [Pereira \(2012\)](#), o operador de mutação na PG não é simples de ser implementado. Primeiramente, é necessário ter certeza de que a árvore do indivíduo se mantém válida após a mutação; isto é, o operador de mutação não pode substituir um nó (ou subárvore) por um nó de tipo de dado diferente. O operador de mutação consiste na troca de um nó (terminal ou não) por outro nó (terminal ou uma subárvore) gerado. É notório que aqui, assim como no cruzamento, os indivíduos podem aumentar ou diminuir de tamanho. A probabilidade do indivíduo sofrer mutação é um parâmetro de entrada da PG ([KOZA, 1992](#)).

Figura 12 – Exemplo de cruzamento na Programação Genética.



Fonte: Adaptado de Koza (1992)

Freitas (2002) classifica a operação de mutação em quatro categorias principais, baseadas na capacidade de alteração de tamanho da árvore:

- **Mutação de Ponto:** Consiste na troca de um nó externo por outro nó externo;
- **Mutação de Colapso:** Um nó externo substitui um nó interno (subárvore). Nesse caso, a árvore diminui sua altura;
- **Mutação de Expansão:** Um nó interno (subárvore) substitui um nó externo. Nesse caso, a árvore aumenta sua altura;
- **Mutação de subárvore:** Um nó interno substitui outro nó interno. Nesse caso, a árvore pode aumentar ou diminuir sua altura ou ainda não ter seu tamanho alterado.

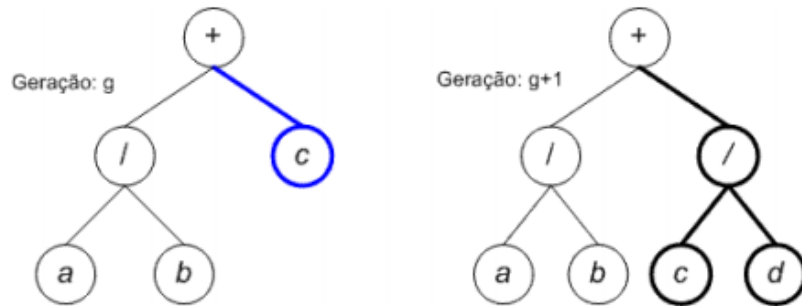
Um exemplo de mutação por expansão pode ser visualizado na Figura 13.

3.3.10 Multi-Classificadores

Sistemas multi-classificadores podem ser definidos como um conjunto de classificadores cujas decisões individuais são combinadas de alguma maneira de modo a classificar novas instâncias (KUNCHEVA, 2004). Entende-se por multi-classificadores: *ensembles*, comitês, *voting*, *mixture of experts* e meta-classificadores.

Estudos como de Dietterich (2000) mostram como multi-classificadores possibilitam um desempenho significativamente superior aos individuais. A ideia da multi-classificação é que

Figura 13 – Exemplo de mutação na Programação Genética.



Fonte: Adaptado de Koza (1992)

nenhum método ou sistema pode ser uniformemente superior a outro, e que a integração de diversos métodos individuais melhora o desempenho do classificador final. Em geral, a utilização de multi-classificadores tem a tendência de diminuir a taxa de erro, tornando o classificador final mais preciso (DŽEROSKI; ŽENKO, 2004). Por esta razão, os estudos de métodos inovadores de multi-classificação cada vez mais têm sido pesquisados (TAN; GILBERT, 2003).

O bom desempenho do sistema baseado em múltiplos classificadores depende dos elementos do conjunto, ou seja, dos componentes do sistema. Um dos problemas citados por Ho (2001) é o desafio em recursos computacionais em termos de treinamento.

Segundo Kuncheva (2004), os múltiplos classificadores podem ser classificados em 4 (quatro) tipos, são eles:

- **Métodos de combinação linear:** as saídas fornecidas pelos classificadores são combinadas linearmente. Como exemplos de métodos de combinação linear têm-se a soma, a média e a mediana;
- **Métodos de combinação não-linear:** as saídas fornecidas pelos classificadores são combinadas de forma não linear, como o que ocorre no uso do voto majoritário e na combinação baseada em ordenação (ranking);
- **Métodos baseados em estatísticas:** utilizam-se probabilidades ou combinações estatísticas nas saídas geradas pelos classificadores;
- **Métodos baseados em inteligência computacional:** técnicas de inteligência computacional (IC) são usadas para combinar as saídas fornecidas pelos classificadores.

Alguns multi-classificadores destacam-se, como: *Bagging* (BREIMAN, 1996), *Boosting* (SCHAPIRE, 1990), *Stacking* (WOLPERT, 1992) e *StackingC* e suas respectivas variações. Na próxima seção será abordado com mais ênfase todos estes métodos.

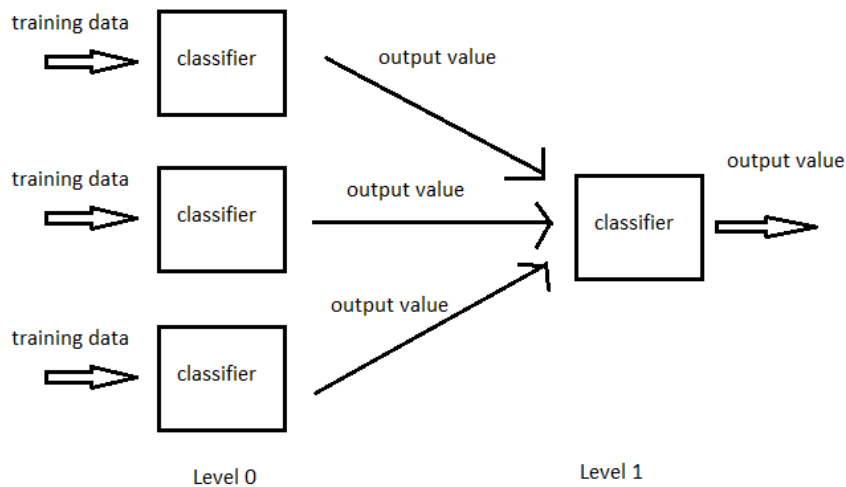
3.3.10.1 Meta-Classificadores

Meta-classificadores são classificadores que incorporam conhecimento sobre o comportamento de classificadores. Eles utilizam as saídas dos classificadores base como entradas para o novo classificador a ser treinado. Alguns métodos da literatura que se destacam são listados a seguir:

- **Bagging (Bootstrap AGGREGATING)** é baseado na amostragem *bootstrap* (KUN-CHEVA, 2004). Este consiste em gerar N subconjuntos de treinamento de forma aleatória sendo que os exemplos selecionados são devolvidos ao conjunto base para geração dos demais subconjuntos. Esta distinção aleatória entre os vários subconjuntos de treinamento confere diversidade aos modelos de classificação que são obtidos a partir de cada um desses subconjuntos. Para combinar os diferentes classificadores base um esquema de votação simples é normalmente empregado. Além disso, o total de classificadores obtidos equivale ao número de subconjuntos gerados, ou seja, N . Outra característica do *Bagging* é que o mesmo algoritmo é aplicado para todos os N subconjuntos, podendo ser gerados em paralelo. O método *Bagging* busca, portanto, classificadores que sejam instáveis, capazes de generalizar de forma distinta, não necessariamente de forma ótima. Isso porque, mesmo que os componentes não apresentem uma boa capacidade de generalização, a combinação deles tendem a generalizar bem.
- **Boosting**: O método possui estratégia semelhante ao *Bagging*. No *Boosting*, o conjunto de treinamento usado por cada classificador base, membro de uma sequência de classificadores, é escolhido baseado no desempenho dos primeiros classificadores. É necessário, assim, que os vários classificadores base do *Boosting* sejam treinados sequencialmente, visando definir as instâncias que irão constituir os próximos conjuntos de treinamento. Conseqüentemente, esta instância terá uma chance maior de ser escolhida para compor o conjunto de treinamento do próximo classificador a ser gerado. Assim, pode-se observar que, ao contrário do que ocorre no *Bagging*, apenas o primeiro classificador base da sequência do método é treinado a partir de uma amostragem aleatória do conjunto de dados original.
- **Stacking**: Também conhecido em inglês como *stacked generalization* (generalização empilhada). O método *Stacking* tenta aprender qual é o classificador mais confiável e qual a melhor maneira para combinar as suas saídas. *Stacking* combina a saída de vários classificadores, que são formados por diferentes algoritmos de aprendizado de máquina, e estas saídas são combinadas por um meta-classificador, o qual gera as saídas com um melhor desempenho. A Figura 14 ilustra a representação deste algoritmo.
- **StackingC**: uma variação do algoritmo *Stacking*, o método *StackingC* (*stacking with confidences*) remove os atributos que não são relevantes e reduz a dimensionalidade

da base de dados antes de utilizá-la como entrada para o meta-classificador.

Figura 14 – Representação do algoritmo *Stacking*.



Fonte: Lo (2017)

Segundo [Zenko, Todorovski e Dzeroski \(2001\)](#), o *Stacking* e sua variação são menos difundidos do que as técnicas de multiclassificação *Bagging* e o *Boosting*. A sua ideia básica pode ser aplicada em muitas variações diferentes. O algoritmo *Stacking* tem dificuldades na definição de qual meta-classificador é mais apropriado, e de quais e quantos classificadores base devem ser utilizados ([WITTEN; FRANK, 2005](#)). [Wolpert \(1992\)](#) denomina esta característica como *black art*.

Outros trabalhos na literatura, como os de [Caffé, Perez e Baranauskas \(2012\)](#), utilizam um número variável de algoritmos base e, para gerar o meta-classificador, utilizam o algoritmo *Naive Bayes* ([HAND; YU, 2001](#)), visto que, além de simples, o *Naive Bayes* é conhecido por ganhar de métodos de classificação altamente sofisticados. [Džeroski e Ženko \(2004\)](#) também recomendam a utilização de algoritmos simples para compor o meta-classificador.

3.4 Métricas de Avaliação

Nesta seção serão apresentadas as métricas de avaliação de desempenho dos modelos de Inteligência Computacional e as métricas de avaliação de risco das estratégias de negociação.

3.4.1 Medidas de Desempenho

As medidas de desempenho têm o objetivo de realizar a avaliação da qualidade dos classificadores (preditores) com base em modelos de aprendizado de máquina, sendo constituídas de fórmulas matemáticas e estatísticas ([ZHANG, 2014](#)). Com as predições

dos classificadores, é possível extrair a matriz de confusão e, assim, gerar as métricas de desempenho e risco.

A matriz de confusão é frequentemente utilizada para organizar e exibir informações utilizadas para avaliar o desempenho de um algoritmo. Cada coluna da matriz representa as instâncias de uma classe prevista, enquanto cada linha representa os casos em uma classe real. Verdadeiros positivos (VP) são exemplos rotulados corretamente como positivos, falsos positivos (FP) referem-se a exemplos negativos incorretamente rotulados como positivos; verdadeiros negativos (VN) correspondem a negativos rotulados corretamente como negativos e falsos negativos (FN) referem-se a exemplos positivos incorretamente rotulados como negativos (DAVIS; GOADRICH, 2006). A Figura 15 ilustra a matriz de confusão para duas classes.

Figura 15 – Matriz de Confusão.

		Valor Previsto		Total
		p	n	
Valor Real	p'	Verdadeiro Positivo	Falso Negativo	P'
	n'	Falso Positivo	Verdadeiro Negativo	N'
Total		P	N	

Fonte: Adaptado de Pereira (2012)

As métricas mais utilizadas são a Acurácia, a Revocação, a Precisão, o *F1-score* e a Especificidade, que são detalhadas abaixo:

- **Acurácia:** é a quantidade de amostras positivas e negativas classificadas corretamente dividida pelo total de amostras da série avaliada em percentual, representada pela Equação 11.

$$Acuracia = \frac{VP + VN}{VP + FP + VN + FN} \quad (11)$$

- **Precisão:** é a quantidade de amostras positivas classificadas corretamente sobre o total de amostras classificadas como falsas positivas acrescida das amostras positivas em percentual, representada pela Equação 12.

$$Precisao = \frac{VP}{VP + FP} \quad (12)$$

- **Revocação:** é a quantidade de amostras positivas (*VP*) classificadas corretamente sobre o total de amostras classificadas como falsas negativas mais *VP* em percentual,

representada pela Equação 13.

$$Revocacao = \frac{VP}{VP + FN} \quad (13)$$

- **F1-score**: é a média harmônica entre a Precisão e a Revocação, representada pela Equação 14.

$$F1 - score = \frac{2 * Precisao * Revocacao}{Precisao + Revocacao} \quad (14)$$

Além das métricas de performance já apresentadas, foram utilizadas também algumas métricas financeiras para a análise dos resultados, visto que em problemas de investimento é importante validar modelos de ponto de vista financeiro, pois estes nem sempre correlacionam-se com a qualidade preditiva do modelo. Assim, entende-se que, por vezes, modelos com precisão baixa são capazes de apresentar lucros superiores, desde que sejam capazes de detectar mudanças mais agudas nos preços.

3.4.2 Medidas Financeiras

As métricas financeiras utilizadas serão: o retorno financeiro, o ganho médio, a perda média, e a quantidade de gatilhos de ganho e gatilhos de perda. Também serão apresentadas algumas medidas de risco, que têm a finalidade de avaliar o risco versus o retorno (ERNST, 2014) (VARGA, 2001) das estratégias de negociação utilizando os seguintes indicadores: volatilidade e índice *Sharpe*. Gatilho é o motivo para execução de origem de compra ou venda, este motivo por exemplo em nossa abordagem pode ser uma tendência de alta ou não alta.

- **Retorno Financeiro**: é o percentual de ganho financeiro de um ativo durante um período, conforme a Equação 1;
- **Ganho Médio**: é o percentual médio de ganho por gatilho;
- **Perda Média**: é o percentual médio de perda por gatilho;
- **Volatilidade**: é a medida da taxa de variação de um ativo num determinado período de tempo, ou seja, é quanto este ativo variou em (%) num determinado período de tempo. Em resumo, é o desvio padrão dos retornos dos ativos, conforme a Equação 15. Onde x é o retorno de cada amostra, \bar{x} é média das amostras e n a quantidade total de amostras.

$$DP = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}} \quad (15)$$

- **Índice Sharpe**: é uma medida de eficiência da relação do risco X retorno, ou seja, quanto maior é o valor do índice *Sharpe*, mais eficiente é o retorno financeiro (RF) encontrado na estratégia de operação. Para o cálculo deste índice, é subtraído do RF

o valor da taxa do CDI² do período avaliado e o resultado é dividido pela volatilidade, conforme a Equação 16.

$$Sharpe = \frac{RF - CDI}{Volatilidade} \quad (16)$$

3.5 Considerações Finais

Neste Capítulo foram apresentados diversos conceitos de extrema importância para a elaboração da modelagem proposta. Foram definidos os conceitos de séries temporais, juntamente com as dificuldades de previsão. Posteriormente, foram explicadas informações sobre o mercado financeiro. Esclareceu-se o método de análise do mercado financeiro que foi utilizado para a tomada de decisão do algoritmo proposto. Por fim, foram apresentadas as técnicas de inteligência computacional em conjunto com as medidas de avaliação que compuseram a solução proposta.

²CDI: Certificado de Depósitos Interbancários é um valor médio dos juros praticados entre os bancos e serve como uma referência para a economia.

Capítulo 4

Metodologia

Este capítulo descreve o método proposto para a previsão de tendências em séries temporais financeiras. Resumidamente, as etapas da metodologia são: [a] extração dos dados, [b] tratamento dos dados, [c] meta-classificação, e [e] análise dos resultados. A Figura 16 ilustra a metodologia proposta.

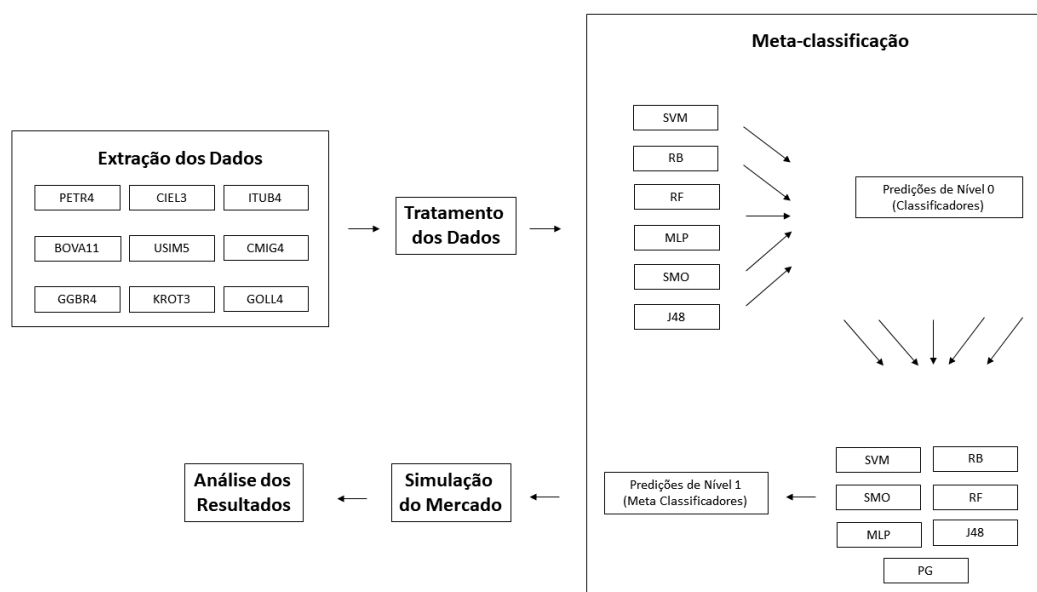


Figura 16 – Metodologia Proposta.

4.1 Extração dos Dados

Para a coleta das séries temporais financeiras foi utilizado o *software MetaTrader*. É um programa criado para realizar negociações *on-line*, fazer análise técnica e utilizar sistemas de *trading* automatizados para os mercados financeiros. O *MetaTrader 5* possui um conjunto de indicadores técnicos que cobrem a maioria das necessidades do *trader*. Também tem acesso a todo o histórico de cotizações do instrumento financeiro (moedas e ações). O

ambiente de desenvolvimento é o *MetaQuotes Language IDE* (MQL). Esse *software* possui diversas funcionalidades e, além disso, é amigável para os desenvolvedores (METATRADER, 2017). As suas vantagens incluem: uma arquitetura orientada a objetos, maior velocidade de cálculo, e uma sintaxe semelhante à da linguagem de programação *C++* (STROUSTRUP, 2000).

Os dados utilizados são oficiais e são arquivos de cotações históricas com dados agrupados por granularidade diária, contendo atributos relativos ao código do ativo, preços (fechamento, abertura, máximo, mínimo, fechamento anterior), quantidade de negócios, quantidade de papéis, volume financeiro, data e hora das transações realizadas de todas as ações negociadas na bolsa de valores brasileira (B3).

4.2 Tratamento dos Dados

A etapa de transformação de dados é responsável por realizar a criação do atributo indicador alvo (classe), e em sequência normalizar cada atributo de entrada: preço de fechamento, abertura, máximo, mínimo, fechamento anterior, quantidade de negócios, quantidade de papéis e volume financeiro extraídos na etapa anterior.

Um indicador alvo (classe) é a variável que procura-se prever, enquanto que um atributo é a variável de entrada, podendo ser mais de um.

Para atribuir atributo indicador alvo *classe*, a política usada para determinar o valor é baseada no preço de fechamento. Onde i é o momento atual e $i + 1$ o seguinte. Desta forma, caso o $fechamento_{(i+1)}$ seja maior que o $fechamento_i$ a classe será 1 (alta), caso contrário 0 (não alta), conforme a Equação 17:

$$classe = \begin{cases} 1 & \text{se } fechamento_{(i+1)} > fechamento_i \\ 0 & \text{se } fechamento_{(i+1)} \leq fechamento_i \end{cases} \quad (17)$$

Após a criação do indicador alvo é realizada a normalização dos dados de entrada. Como apresentado no capítulo 3, a maioria dos estudos financeiros, inclusive esta tese, ao invés de utilizarem os preços absolutos dos ativos, optam por utilizar os retornos gerados pelos mesmos devido aos inúmeros benefícios já apresentados. Sendo o principal motivo que modelos estatísticos têm dificuldades em operar em séries cuja componente de tendência é acentuada. Dessa forma, o objetivo desta normalização é reduzir esta tendência.

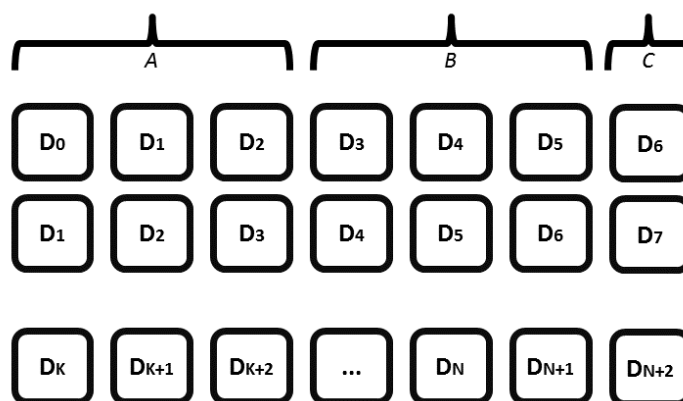
4.3 Meta-classificação

A fim de prever a tendência dos retornos das ações, foi desenvolvido um algoritmo de meta-classificação que, em seu conjunto de classificadores base (nível 0), possui 6 técnicas de aprendizado de máquina: Florestas Aleatórias (RF), Máquinas de Vetor de Suporte (SVM), Perceptron Multicamadas (MLP), Árvore de Decisão (J48), Otimização Mínima Sequencial (SMO) e Redes Bayesianas (BN), onde as entradas dos classificadores base são os retornos normalizados com granularidade diária. Já em seu conjunto de meta-classificadores (nível 1), foram utilizadas as 6 técnicas já citadas, mais a Programação Genética (PG). Segundo [Bordin \(2018\)](#) a utilização da PG em conjunto com outras técnicas de aprendizado de máquina, organizados em um comitê de classificadores, pode incrementar o poder de predição e o resultado geral do sistema, resultando em valores superiores aos alcançados individualmente pelas técnicas.

As séries temporais de financeiras constantemente seu comportamento sofre mudanças e são influenciadas por muitos fatores externos de toda espécie (políticos, locais, ou macroeconômicos), e por isso modelos capazes de generalizar no longo prazo são algo distante de se conseguir, logo prima-se pelo uso de apenas dados recentes.

Para este trabalho utilizamos o conceito de janela deslizante, no qual a massa de dados de treinamento e teste se move conforme o tempo. Mais especificamente, um modelo novo de predição é gerado a cada novo período de negociação. A Figura 17 mostra um exemplo da janela deslizante.

Figura 17 – Exemplo de Janela Deslizante.



Fonte: Elaborada pelo autor

Seguindo o exemplo da Figura 17, parametrizando uma janela deslizante de 7 (sete) dias. O tamanho da janela N é fixado e a janela se move para a direita um dia de cada vez, enquanto um dia do lado esquerdo é descartado. Para cada deslocamento da janela, um novo modelo é gerado para a previsão do dia seguinte.

Assim neste exemplo, a primeira configuração do experimento usaria o conjunto de treino A que possui os dias D_0, D_1 e D_2 para treinar os classificadores de forma isolada; os dias D_3, D_4, D_5 como conjunto de treino B para os meta-classificadores; e o conjunto de teste C com o dia D_6 . Seguindo esta lógica, a próxima janela se inicia no dia D_1 , visto que o conjunto de testes neste exemplo é somente de um único dia. Assim, o número de janelas finalizaria até que não existissem mais dias no conjunto de testes.

A abordagem de meta-classificação proposta é apresentada no Algoritmo 21. Mais especificamente o Algoritmo foi construído em duas fases. Na primeira fase, no nível 0, um conjunto de classificadores base é gerado por diferentes algoritmos de aprendizado de máquina e um único conjunto de dados de treinamento. Na segunda fase, nível 1, os meta-classificadores combinam as saídas dos classificadores base em uma predição final. Os dados dos meta-classificadores tem, pois, tantos atributos quantos forem os classificadores base.

Os parâmetros de entrada do arcabouço são: o ativo a ser avaliado, o conjunto de classificadores base (nível 0), o conjunto meta-classificadores (nível 1), por fim o tamanho dos conjuntos de dados de treino nível 0, treino nível 1 e testes, estes estão representados pelas variáveis *ativo*, *classificadores* *metaClassificadores*, s_a , s_b e s_c .

Nas linhas 1 e 2 do algoritmo, são realizadas as etapas de extração dos dados 4.1 e tratamento dos dados 4.2, conforme detalhe em suas respectivas seções. Nas linhas 3, 4 e 5 são armazenados o tamanho correspondente a cada conjuntos de dados. Nas linhas 9, 10 e 11 os dados são divididos, respeitando o número de instâncias previamente armazenado.

Com o objetivo de gerar os modelos dos os classificadores base, treina-se cada classificador com o conjunto A , assim que todos os modelos tenham sido gerados (linha 12), aplicam-se os conjuntos B e C nos modelos, gerando assim dois novos conjuntos B' e C' (linhas 13 e 14); esses conjuntos serão compostos de previsões de cada técnica, gerando 2 matrizes binárias, onde 1 é sinal de alta e 0 sinal de não alta, esses são os meta-atributos gerados.

O próximo passo (linha 12) é treinar os meta-classificadores (nível 1), utilizando o conjunto B' e testando no conjunto C' , para obtenção das previsões dos meta-classificadores (linha 16). Sendo assim, estas previsões *logPredicoes* serão utilizadas na próxima etapa de

simulação do mercado.

Algoritmo 1: PSEUDO-CÓDIGO DO META-CLASSIFICADOR

Entrada: *ativo*, *classificadores*, *metaClassificadores*, s_a , s_b , s_c

Saída: *logPredicoes*

```
1 dados ← obterDados(ativo);
2 dadosTratados ← tratamentoDosDados(dados);
3 A.setTamanhoDoConjunto( $s_a$ );
4 B.setTamanhoDoConjunto( $s_b$ );
5 C.setTamanhoDoConjunto( $s_c$ );
6 início
7   para  $j < dadosNomalizados.getQuantidade()$  faça
8     para  $i < classificadores.size()$  faça
9       A ← dadosTratados.get( $j$ , A.size());
10      B ← dadosTratados.get( $j + A.size()$ , B.size());
11      C ← dadosTratados.get( $j + A.size() + B.size()$ , C.size());
12      modelo ← treinar(classificadores[ $i$ ], A);
13      B' ← classificar (modelo, B);
14      C' ← classificar (modelo, C);
15      metaModelo ← treinar(metaClassificadores[ $i$ ], B');
16      logPredicoes ← classificar (metaModelo, C');
17     fim
18      $j \leftarrow j + C.size()$ ;
19   fim
20 fim
21 retorna logPredicoes
```

4.4 Simulação do Mercado

Foi desenvolvido um módulo capaz de avaliar a qualidade do uso das estratégias de negociação. É importante destacar que uma característica do sistema de negociação desenvolvido é a granularidade, ou seja, *candles* diários. Esta decisão foi tomada devido à alta liquidez da bolsa de valores no fechamento do dia. Criamos sete estratégias de negociação com a ideia base de comprar com previsão de alta e vender com previsão de não alta nos seguintes cenários:

1. sem *stops*;
2. com *stop loss*;
3. com *stop gain*;
4. com *stop loss* e *stop gain*;

5. com *stop* de período;
6. com *stop loss*, *stop gain* e *stop* de período;
7. compra com previsão de alta e vende no próximo período independente do sinal;

A funcionalidade de *stop* é uma medida de segurança que o investidor possui ao operar na Bolsa de Valores; e têm como função definir a perda máxima tolerada (*stop loss*) e o lucro almejado (*stop gain*). O *stop loss* é uma funcionalidade que permite ao investidor configurar uma venda automática de determinadas ações caso elas atinjam um valor ajustado por ele. Isso serve principalmente para limitar o prejuízo de uma operação, mas também para manter parte do lucro obtido – caso haja a reconfiguração do *stop loss* depois da valorização dos ativos. O *stop gain* é também uma configuração para a venda automática, mas na situação contrária: quando a ação alcançar um determinado valor, é vendida. A motivação é simples: estas operações têm como objetivo o controle dos riscos (INVESTIDOR, 2017). Já o *stop* por período foi um conceito que criamos para garantir a venda em um determinado período sem operações.

Para uma maior clareza do funcionamento básico das abordagens acima, segue o pseudo-código da estratégia “sem *stops*” no Algoritmo 21.

Os parâmetros de entrada do algoritmo de simulação são todas as previsões realizadas na etapa anterior, ativo a ser avaliado e o valor do investimento inicial, estes estão representados pelas variáveis *ativo*, *logPredicoes* e *investimentoInicial*. Na linha 1 a variável *posicao* realiza o controle de posição (comprado ou vendido), visto que nesta proposta, quando é executada uma ordem de compra, utiliza-se de todo o capital e quando vende-se, realiza-se todo lote de ações.

Sendo assim, para cada previsão realizada é avaliada sua tendência (linha 6), caso a tendência seja 1 (alta), realiza-se uma operação de compra, caso contrário uma operação de venda. Ao final da avaliação da estratégia de operação proposta, para cada ativo, serão armazenados os dados e parâmetros que fazem parte do processo de execução das ordens de compra e venda (gatilhos) como, por exemplo, a quantidade de gatilhos e o percentual

(%) (ganho, perda e retorno) acumulados no arquivo contendo o *log* de operação.

Algoritmo 2: PSEUDO-CÓDIGO DA ESTRATÉGIA “SEM stops”

Entrada: *ativo, logPredicoes, investimentoInicial*

Saída: *logOperacao*

```
1 posicao ← vendido;
2 dados ← obterDados(ativo);
3 início
4   para i < logPredicoes.size() faça
5     tendencia ← obterTendencia(logPredicoes.get(i));
6     se posicao = vendido então
7       se tendencia = 1 então
8         comprar (dados[i].getPrecoFechamento());
9         posicao ← comprado;
10      fim
11     fim
12     senão
13       se tendencia = 0 então
14         vender (dados[i].getPrecoFechamento());
15         posicao ← vendido;
16       fim
17     fim
18     logOperacao ← calcular(investimentoInicial);
19   fim
20 fim
21 retorna logOperacao
```

4.5 Análise dos Resultados

A etapa da análise de resultados apresenta três tipos de análises distintas consolidadas para cada código de ativo avaliado; as análises: estatística, de desempenho, e de operação/risco.

A análise estatística apresenta a caracterização da amostra de cada código de ativo e é composta pela distribuição dos dados em relação aos períodos de treino e teste da série temporal avaliada, contemplando os seguintes indicadores:

- Quantidades de dias avaliados da série;
- Preço Inicial do código de ativo;
- Preço Final do código de ativo; e
- Percentual de dias em que ocorreram altas e não altas.

A análise de desempenho avalia os modelos de classificação através das métricas previamente descritas no Capítulo 3 com a finalidade de medir as taxas de desempenho dos modelos de previsão (classificadores) durante a simulação realizada:

- **Acurácia** (Equação 11);
- **Precisão** (Equação 12);
- **Revocação** (Equação 13);
- **F1-score** (Equação 14); e

Para análise financeira foram utilizadas as seguintes métricas:

- **Percentuais (%) de:** Dias Positivos, Dias Negativos, Perda, Ganho e Retorno Total;
- **Índice Sharpe:** avalia a rentabilidade e o risco de um investimento, sendo fundamental para mensurar o quanto de retorno excedente em relação a um ativo livre de risco é compensado através de seu nível de risco. Conforme a Equação 16
- **Drawdown:** é o percentual de queda a partir do último ponto máximo de retorno financeiro cumulativo, *drawdown* máximo é o percentual máximo de queda em uma determinada série de dados.

Todas as métricas, sejam elas de performance ou financeiras, são comparadas entre os classificadores isolados, os meta-classificadores, alguns *baselines* e *benchmarks* do período.

Dentre os *baselines* estão:

- **Buy-and-hold:** assumir posição de compra no período inicial t_1 e fechar a posição (vender) no período final t_n .
- **Random:** modelo que sorteia a operação de compra ou venda aleatoriamente.
- **Estratégia Inversa:** operar realizando a ação contrária ao que o modelo predictor indica.

Já os *benchmarks* para o mercado brasileiro, são:

- **Certificados de Depósitos Interbancários (CDI):** é uma média dos juros praticados entre os Bancos e serve como uma referência para o preço do dinheiro na economia, pois é utilizada como *benchmark* em muitos investimentos (FINANCIAL, 2017);
- **IGP-M:** Índice Geral de Preços do Mercado é o indicador de movimento dos preços calculado mensalmente pela FGV (Fundação Getúlio Vargas) e divulgado no final de cada mês de referência, sendo utilizado para o reajuste dos aumentos da energia elétrica e dos contratos de aluguéis (ADVFN, 2010);
- **Cotação do Dólar em Reais:** o dólar é a moeda oficial dos Estados Unidos e utilizada no mundo inteiro, tanto em reservas internacionais, como em livre circulação em

alguns países;

- **IBOV:** O Ibovespa é o principal indicador de desempenho das ações negociadas na B3 e reúne as empresas mais importantes do mercado de capitais brasileiro. Foi criado em 1968 e, ao longo desses 50 anos, consolidou-se como referência para investidores ao redor do mundo.

O propósito desta etapa é avaliar a viabilidade e o grau de risco de investimento das propostas de estratégias de operação utilizando algoritmos de aprendizado de máquina combinados e a possibilidade de aplicação futura destas estratégias em um cenário real no mercado financeiro.

Para a validação da metodologia proposta, todo o arcabouço foi parametrizado para uma quantidade de 30 execuções.

4.6 Considerações Finais

Este capítulo apresentou a proposta de uma metodologia para a previsão de tendências em séries temporais financeiras. Demonstrando cada etapa de forma detalhada: a coleta dos conjuntos de dados e seus atributos, a etapa de transformação de dados, a apresentação do algoritmo de meta-classificação construído, além das estratégias de operações desenvolvidas e, por último, a descrição das formas de análise de resultados. O próximo capítulo apresentará toda a instanciação da metodologia desenvolvida, ou seja, os ativos e os períodos utilizados, a parametrização dos algoritmos de aprendizado de máquina e o tamanho de janelas. Por fim, os resultados obtidos.

Capítulo 5

Experimentos e Resultados

Como visto anteriormente, neste trabalho investiga-se o uso de meta-classificadores para a previsão de séries temporais financeiras. Neste capítulo serão apresentados os conjuntos de dados, parametrizações utilizadas na instanciação dos testes e, por fim, os resultados. Para a aplicação e validação da metodologia, foram utilizados 9 conjuntos de dados históricos da B3.

Os códigos de ativos que foram analisados nos experimentos são de empresas que compõem o Índice Bovespa (Ibovespa). Esses ativos foram escolhidos por possuírem uma maior representatividade em relação ao volume negociado, representando diferentes setores da economia como petróleo e gás, mineração, bancos, siderurgia, aviação e educação, listados a seguir:

- **Petrobrás (PETR4)**: ações da Petróleo Brasileiro S.A. Esta é uma empresa de capital aberto, cujo acionista majoritário é o Governo do Brasil, sendo, portanto, uma empresa estatal de economia mista¹;
- **Cielo (CIEL3)**: ações da empresa Cielo (antiga Visanet Brasil). Esta é uma empresa brasileira que atua como adquirente multi-bandeira, sendo uma das responsáveis pela captura, transmissão e liquidação financeira de transações com cartões de crédito e débito²;
- **Itaú Unibanco (ITUB4)** ações da empresa Itaú Unibanco. Este é um banco brasileiro fundado em 4 de novembro de 2008 mediante a fusão de duas das maiores instituições financeiras do país, o *Holding Itaú Bancos e Unibanco*³;
- **Índice Bovespa (BOVA11)**: acompanha o Índice Bovespa. O objetivo desse fundo é obter uma performance próxima à do próprio Índice Bovespa (IBOV). Dessa forma, vários investidores unidos investem nesse fundo, que utiliza todo o recurso reunido para comprar ações exatamente nas mesmas proporções do IBOV;

¹<http://www.petrobras.com.br/>

²<http://www.cielo.com.br/>

³<http://www.itau.com.br/>

- **Usiminas (USIM5)**: uma empresa do setor siderúrgico líder na produção e comercialização de aços planos laminados a frio e a quente, bobinas, placas e revestidos, destinados principalmente aos setores de bens de capital e de bens de consumo da linha branca, além da indústria automotiva⁴;
- **Companhia Energética de Minas Gerais (CMIG4)**: é uma das principais concessionárias de energia elétrica do Brasil, tendo sede na cidade de Belo Horizonte, capital do estado de Minas Gerais⁵;
- **Gerdau (GGBR4)**: também referida como Grupo Gerdau, é uma empresa siderúrgica brasileira regida sob a forma jurídica de sociedade anônima. Possui operações industriais em 11 países – nas Américas, na Europa e na Ásia, as quais somam uma capacidade instalada superior a 25 milhões de toneladas de aço bruto por ano⁶;
- **Kroton Educacional (KROT3)**: é a maior empresa privada do mundo no ramo da educação. É a maior empresa brasileira no segmento de ensino superior para o número de alunos e de receita⁷;
- **Gol Linhas Aéreas Inteligentes (GOLL4)**: é a maior companhia aérea do Brasil em número de passageiros, tendo 36% de participação do mercado doméstico⁸.

Além dos experimentos que serão mencionados neste Capítulo, foram realizados diversos outros, utilizando configurações e parametrizações diferentes nos algoritmos com o intuito de testar diferentes variantes até chegar ao modelo atual, que é resultado da melhor configuração encontrada.

5.1 Caracterização e Análise dos Dados

Foram coletados dados de *candles* diários de agosto de 2015 a agosto de 2016. A Tabela 1 detalha informações de preço dos ativos utilizados. A Figura 18 mostra a evolução dos preços durante este período. Em geral, as séries de preços apresentam uma alta volatilidade. É interessante observar que em quase todos ativos entre o 100º dia e o 150º dia da série, houve pontos de menor preço. Segundo Almeida (2016), este fato aconteceu devido ao caos político no país, a economia despencou, fazendo o país perder a confiança dos investidores nacionais e internacionais.

Avaliando a proporção das classes (Figura 19), é possível verificar que a distribuição é bastante balanceada para os ativos selecionados. Na maioria dos casos, o número de instâncias de classe de alta é ligeiramente superior ao de não alta.

⁴<http://www.usiminas.com.br/>

⁵<http://www.cemig.com.br/>

⁶<http://www.gerdau.com.br/>

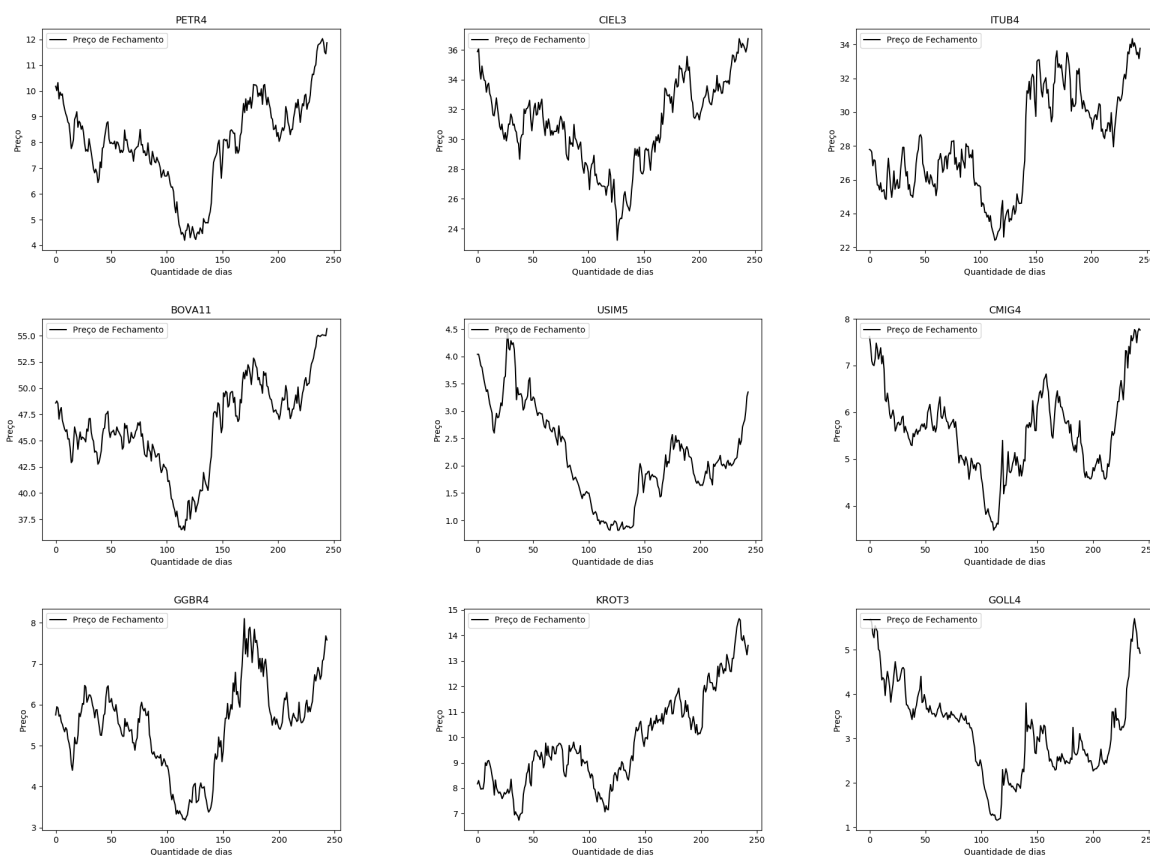
⁷<http://www.kroton.com.br/>

⁸<http://www.voegol.com.br/>

Tabela 1 – Valor de Mercado de Ativos entre Agosto de 2015 e Agosto de 2016.

Ativo	Quantidade de Dias	Preço Inicial	Preço Final	% Não Alta	% Alta
PETR4	244	R\$ 10,18	R\$ 11,87	47%	53%
CIEL3	244	R\$ 35,88	R\$ 36,75	50%	50%
ITUB4	244	R\$ 27,79	R\$ 33,78	48%	52%
BOVA11	244	R\$ 48,59	R\$ 55,68	49%	51%
USIM5	244	R\$ 4,04	R\$ 3,35	53%	47%
CMIG4	244	R\$ 7,62	R\$ 7,76	50%	50%
GGBR4	244	R\$ 5,75	R\$ 7,58	49%	51%
KROT3	244	R\$ 8,17	R\$ 13,61	49%	51%
GOLL4	244	R\$ 5,67	R\$ 4,92	58%	42%

Figura 18 – Série de Preços de Fechamento Original.

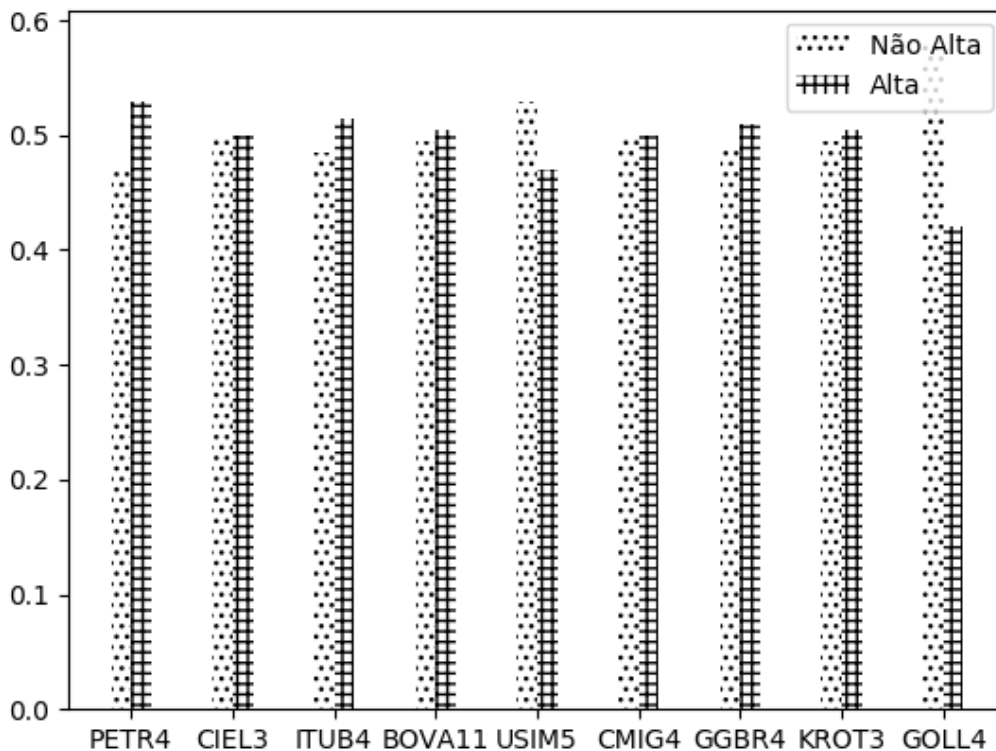


Com intuito de comparar os resultados financeiros, também foram coletados alguns indicadores de mercado (*benchmarks*). A Tabela 2 apresenta os *benchmarks* do período avaliado, contendo: CDI, IGP-M, Dólar, Ouro e Índice Bovespa (IBOV).

Tabela 2 – *Benchmarks*

CDI	IGP-M	Dólar	Ouro	IBOV
19,27%	11,60%	29,70%	17,83%	14,76%

Figura 19 – Distribuição das classes por ativo.



5.2 Instanciação dos Algoritmos

Para executar a solução proposta, foi necessário instanciar todos os classificadores e o meta-classificador. Os algoritmos SVM, RF, MLP, J48, SMO, BN e PG foram parametrizados conforme detalhamento nas respectivas Tabelas 3, 4, 5, 6, 7 e 8.

Tabela 3 – Parâmetros SVM.

SVM		
Parâmetro	Descrição	Conteúdo
kernel	Função kernel	Radial Basis Function
gamma	Valor do parâmetro do kernel RBF	0,25
cost	Parâmetro de penalidade	1

Tabela 4 – Parâmetros RF.

RF		
Parâmetro	Descrição	Conteúdo
number_attribute	Número de atributos utilizados	$\text{int}(\log_2(\text{atributos})+1)$
numTrees	Número de arvores da floresta	100
depth	Profundidade máxima da árvore	ilimitado

Tabela 5 – Parâmetros MLP.

MLP		
Parâmetro	Descrição	Conteúdo
epochs	Parâmetro de parada do processo	500
momentum	Termo de Momentum	0,2
learningRate	Taxa de aprendizagem	0,3
hiddenLayers	Número de camadas ocultas	(atributos + classes) / 2

Tabela 6 – Parâmetros J48.

J48		
Parâmetro	Descrição	Conteúdo
confidenceFactor	Factor de confiança utilizado na poda	0,25
minNumObj	Número mínimo de instâncias por folha	2

Tabela 7 – Parâmetros SMO.

SMO		
Parâmetro	Descrição	Conteúdo
kernel	Função kernel	Radial Basis Function
gamma	Valor do parâmetro do kernel RBF	0,01
tolerancia	Tolerância	0,001
exponent	Expoente do kernel polinomial	1.0E-12
c	Parâmetro C de complexidade	1

Tabela 8 – Parâmetros PG.

PG		
Parâmetro	Descrição	Conteúdo
tipo_mutacao	Tipo de mutação	Um ponto
tipo_cruzamento	Tipo de cruzamento	Um ponto
taxa_mutacao	Taxa de mutação	0,15
taxa_cruzamento	Taxa de cruzamento	0,85
num_geracoes	Número de gerações	200
fitness	Função de fitness	Acurácia
selecao	Tipo de seleção	torneio
tam_populacao	Tamanho da população	500
oper	Tipos dos operadores	AND

Após várias simulações com alternância de diversos parâmetros com o objetivo de melhorar a performance do arcabouço aos dados em questão, os mesmos foram calibrados de acordo com a Tabela 9, na qual a coluna “Resultado” apresenta a configuração final utilizada nos experimentos. Foram realizadas 30 execuções por ativo.

Em resumo, o meta-classificador foi inicializado com 7 algoritmos já citados. Os conjuntos de treinamento e testes foram divididos em três sub-conjuntos em uma janela deslizante: o conjunto *A* possui tamanho de 20 *candles*; o conjunto *B* possui 10; e o conjunto *C*, 10 também.

Para as estratégias de negociação, aquelas que possuem *stops*, foi utilizado 10 dias para o *stop* de período, juntamente com percentuais 2% e 3% para *loss* e *gain*, respectivamente.

Tabela 9 – Intervalo de parâmetros de calibração do Meta-Classificador

Parâmetros	Varição	Resultado
Quantidade de conjuntos	[2;3]	3
Conjunto A	[5; 50]	20
Conjunto B	[5; 50]	10
Conjunto C	[1; 30]	10
Stop de Período	[5; 15]	10%
Stop Loss	[1; 5]	2%
Stop Gain	[1; 5]	3%
Estratégia	[1; 7]	1

5.3 Avaliação dos Modelos e Estratégias

Diante do grande volume de informação gerado pelo arcabouço, é muito importante selecionar os resultados de quais modelos e quais estratégias apresentar. Desta forma, após a calibração dos parâmetros do meta-classificador e a execução do mesmo, os resultados gerados como saída do arcabouço são a combinação dos seguintes dados:

- *ativos*: 9 Ativos da B3;
- *modelos*: 13 Modelos Preditivos (6 Classificadores Isolados + 7 Meta-Classificadores);
- *estrategias*: 7 Estratégias de Simulação do Mercado;
- *medidas*: 11 Medidas de Desempenho.

Assim, através da Equação 18, é possível apresentar a quantidade total de indicadores que são gerados para a avaliação dos resultados. Neste caso, são gerados 9009 indicadores. O que torna a análise impraticável caso não sejam definidos critérios e filtros de avaliação.

$$indicadores = ativos \times modelos \times estrategias \times medidas \quad (18)$$

Sendo assim, com a finalidade de filtrar e diminuir o espaço de exploração de possibilidades de combinações com os modelos de previsões de tendência propostos na apresentação dos resultados, foram definidos 4 filtros macros, são eles:

1. **Estratégia:** o objetivo deste filtro é utilizar a melhor estratégia de simulação do mercado do arcabouço.
2. **Percentual Médio de Retorno Financeiro:** entre todas as informações de um investimento, uma das mais relevantes para qualquer investidor é o Retorno Financeiro; este filtro tem como objetivo selecionar os modelos que possuem percentual positivo médio de retorno para todos ativos e com retornos financeiros superiores ao CDI, IGP-M, Dólar, Ouro e IBOV, conforme a Tabela 2;
3. **Acurácia Média:** este filtro permite selecionar os modelos que apresentam resultados de Acurácia acima dos 50%. Esta métrica foi utilizada como filtro, pois as estratégias de simulação apresentadas utilizam os sinais de compra e venda dos modelos, ou seja, se faz necessário utilizar a predição tanto de “alta” quanto de “não alta”. Além disso, os conjuntos de dados utilizados são balanceados, o que permite utilizar esta medida como desempenho geral do modelo.
4. **Percentual Médio de Retorno Financeiro entre o melhor classificador isolado e o melhor meta-classificador:** O intuito deste critério é demonstrar e selecionar o melhor classificador isolado e o melhor meta-classificador com o objetivo de explorar e aprofundar com mais detalhes apenas os resultados deste filtro.

A Tabela 10 apresenta os resultados simulados de Retorno Financeiro de todos os modelos para todos os ativos. Os resultados foram agrupados por classificador com o intuito de dar facilidade de visualização aos resultados. Desta forma, é possível visualizar que cada classificador possui também um meta-classificador, com exceção da Programação Genética, cujos resultados podem ser visualizados na Tabela 12.

Com o objetivo de facilitar também a identificação do melhor modelo para cada ativo entre os grupos, as colunas com melhor resultado por ativo foi destacado o texto em negrito; e para identificar o melhor modelo entre todos os ativos, a coluna com o melhor modelo foi sublinhada. A Tabela 11 apresenta os resultados de Acurácia com as mesmas características de visualização.

Ao verificar individualmente os resultados de cada ativo, é visivelmente clara certa superioridade em retornos financeiros dos meta-classificadores em grande parte dos ativos. Outro resultado interessante é que os meta-classificadores são superiores em todos os retornos financeiros médios. Também é possível visualizar o mesmo comportamento na acurácia.

Após a apresentação destes resultados, foi realizada cada etapa dos filtros elaborados. O primeiro filtro foi definido durante a etapa de calibragem do arcabouço, assim identificou-se

que, apesar de sua simplicidade, a Estratégia 1 (sem *stops*) demonstrou melhores retornos financeiros em todos os ativos quando comparada às demais estratégias. O segundo filtro mostrou-se bem arrojado e proporcionou uma seleção bem rigorosa; foram selecionados somente os modelos com percentuais de retorno financeiros positivos e com retornos acima dos índices de mercado (CDI, IGP-M, Ouro, Dólar), e foram selecionados todos os meta-classificadores e apenas um classificador isolado, no caso, a MLP com retorno financeiro médio de 30,87% no período. Ao aplicar o quarto filtro, temos como resultado os melhores modelos de cada grupo. Representando os classificadores isolados, selecionou-se a MLP; e representando os meta-classificadores, o Meta-RF obteve melhor performance. A Tabela 13 compara os resultados desse último filtro.

Desta forma, com a execução dos filtros, o objetivo de trabalhar com um conjunto de resultados reduzido foi cumprido, sem perder informação e qualidade para uma análise mais criteriosa dos resultados. Nas próximas seções serão apresentados de forma mais substancial e com mais profundidade os resultados desses dois modelos selecionados.

Tabela 10 – Comparação de Retorno Financeiro Acumulado entre pares de Classificadores e Meta-Classificadores.

Algoritmo	BOVA11	CIEL3	ITUB4	PETR4	USIM5	CMIG4	GGBR4	KROT3	GOLL4	Média
SVM	-6,42%	7,69%	27,31%	-9,82%	14,53%	12,19%	29,00%	47,52%	-3,03%	13,22%
Meta-SVM	58,89%	35,41%	45,74%	28,47%	74,23%	49,63%	61,41%	37,97%	67,28%	<u>51,00%</u>
RF	-5,14%	-1,60%	15,17%	13,24%	71,17%	33,82%	32,19%	84,54%	-12,15%	25,70%
Meta-RF	29,24%	41,23%	56,42%	55,49%	139,84%	86,38%	50,22%	60,93%	57,24%	<u>64,11%</u>
MLP	3,78%	-12,02%	7,29%	21,58%	79,01%	0,84%	17,55%	64,78%	71,13%	28,21%
Meta-MLP	18,67%	30,77%	37,28%	48,37%	90,77%	31,38%	-3,90%	23,07%	37,22%	<u>34,85%</u>
J48	-2,78%	-26,92%	27,15%	42,79%	74,09%	53,46%	9,30%	10,78%	17,00%	22,77%
Meta-J48	19,85%	25,98%	47,74%	17,43%	138,73%	9,04%	34,23%	26,10%	128,42%	<u>49,73%</u>
SMO	5,57%	9,36%	34,81%	16,51%	19,98%	15,17%	26,07%	25,98%	-5,85%	16,40%
Meta-SMO	22,24%	34,28%	32,00%	21,13%	165,00%	55,62%	13,89%	95,32%	79,99%	<u>57,72%</u>
BN	-11,18%	-6,22%	42,16%	-5,57%	35,80%	-6,79%	37,78%	47,20%	15,65%	16,54%
Meta-BN	17,58%	55,06%	19,65%	11,52%	48,74%	17,38%	33,97%	28,60%	70,06%	<u>33,62%</u>

5.4 Avaliação da Viabilidade dos Custos Operacionais

Durante a execução de qualquer simulação de estratégias de negociação, um ponto muito importante são os custos operacionais. Um modelo de previsão pode até ter retornos financeiros positivos, porém, caso o custo operacional seja maior do que o ganho bruto, toda uma estratégia de operação pode ser inviabilizada.

Tabela 11 – Comparação da Acurácia entre pares de Classificadores e Meta-Classificadores.

Algoritmos	BOVA11	CIEL3	ITUB4	PETRA4	USIM5	CMIG4	GGBR4	KROT3	GOLL4	Média
SVM	47,50%	45,00%	45,50%	48,50%	52,00%	49,00%	51,50%	53,00%	56,50%	49,83%
Meta-SVM	56,00%	54,00%	51,50%	50,50%	56,50%	53,50%	54,00%	56,00%	57,00%	<u>54,33%</u>
RF	44,00%	43,00%	50,50%	55,50%	57,50%	47,50%	53,00%	54,00%	53,50%	50,94%
Meta-RF	56,50%	53,00%	53,50%	48,00%	56,50%	54,00%	53,50%	56,00%	53,50%	<u>53,83%</u>
MLP	45,00%	44,50%	51,00%	54,50%	55,00%	49,00%	50,00%	56,50%	52,00%	50,83%
Meta-MLP	50,50%	51,00%	51,50%	51,00%	54,50%	53,50%	50,00%	52,00%	52,00%	<u>51,78%</u>
J48	44,50%	42,50%	51,50%	49,50%	51,00%	46,00%	47,00%	47,50%	55,00%	48,28%
Meta-J48	49,50%	50,50%	53,50%	47,50%	55,00%	50,00%	52,00%	54,50%	58,00%	<u>52,28%</u>
SMO	49,50%	48,50%	50,00%	54,50%	52,50%	50,00%	51,00%	49,50%	52,00%	50,83%
Meta-SMO	49,00%	52,00%	50,00%	56,50%	58,00%	54,00%	50,00%	56,50%	54,50%	<u>53,39%</u>
BN	45,00%	44,50%	50,50%	50,50%	51,50%	51,00%	52,50%	51,50%	57,00%	50,44%
Meta-BN	47,00%	55,00%	49,00%	55,50%	56,00%	47,00%	49,50%	53,50%	57,00%	<u>52,17%</u>

Tabela 12 – Comparação de Retorno Financeiro entre melhor Classificador com Meta-GP.

Algoritmo	BOVA11	CIEL3	ITUB4	PETRA4	USIM5	CMIG4	GGBR4	KROT3	GOLL4	Média
MLP	3,78%	-12,02%	7,29%	21,58%	79,01%	0,84%	17,55%	64,78%	71,13%	28,21%
Meta-GP	36,18%	23,04%	32,61%	42,77%	51,20%	-0,07%	37,03%	52,76%	141,84%	<u>46,37%</u>

Tabela 13 – Comparação de Retorno Financeiro entre melhor Classificador com Meta-Classificador.

Algoritmo	BOVA11	CIEL3	ITUB4	PETRA4	USIM5	CMIG4	GGBR4	KROT3	GOLL4	Média
MLP	3,78%	-12,02%	7,29%	21,58%	79,01%	0,84%	17,55%	64,78%	71,13%	28,21%
Meta-RF	29,24%	41,23%	56,42%	55,49%	139,84%	86,38%	50,22%	60,93%	57,24%	<u>64,11%</u>

Assim, para todas as simulações realizadas neste trabalho foram contabilizados os custos das operações, contemplando o Imposto de Renda (IR) recolhido na fonte pela Bovespa, o Imposto de Renda recolhido pelo investidor sobre o lucro, a Taxa de Corretagem sobre as ordens de compra e venda, o Imposto Sobre Serviços (ISS) sobre o valor da taxa de corretagem e os Emolumentos sobre o valor investido. Devido ao fato de muitas corretoras isentarem seus clientes da taxa de custódia, esta não foi acrescentada nos custos das transações.

A Tabela 14 apresenta os valores dos custos operacionais de forma detalhada durante os experimentos acrescidos de 3 exemplos de lucro bruto (R\$100,00, R\$1.000,00 e R\$10.000,00), o que torna evidente o fato de que, quanto maior for o valor investido, menor será o custo operacional relativo, com exceção da tributação do Imposto de Renda que é um valor fixo de 20,00% sobre o ganho de capital para transações de compra e venda no mesmo dia e

15% sobre os lucros de operações de mais de um dia.

Segundo o site da B3, no Brasil atualmente existem 87 corretoras ativas. Contudo, cada uma utiliza uma taxa de corretagem diferente e também existem corretoras que não praticam a cobrança destas taxas. Oliveira (2019) apresenta uma recente comparação entre as principais corretoras brasileiras. Desta forma, como não existe uma padronização dos custos operacionais, optou-se por utilizar o custo operacional de 20,59% apresentado na Tabela 14, visto que o capital inicial utilizado das simulações foi de R\$10.000,00. Entende-se desta forma que o arcabouço demonstra certa compatibilidade com o mercado real sobre os custos para este cenário, com uma margem de erro prevista.

Tabela 14 – Exemplos de Custos Operacionais.

	Lucro bruto		
	R\$ 100,00	R\$ 1.000,00	R\$ 10.000,00
IR na fonte (0,50%)	R\$ 0,50	R\$ 5,00	R\$ 50,00
IR recolhido pelo investidor (20,00%)	R\$ 20,00	R\$ 200,00	R\$ 2.000,00
Corretagem por ordem (2,50)	R\$ 5,00	R\$ 5,00	R\$ 5,00
ISS sobre a corretagem (12,00%)	R\$ 0,60	R\$ 1,20	R\$ 1,20
Emolumentos (0,025%)	R\$ 0,03	R\$ 0,25	R\$ 2,50
Custo da operação (R\$)	R\$ 26,13	R\$ 211,45	R\$ 2.058,70
Custo da operação (%)	26,13%	21,15%	20,59%
Custo da operação sem IR (%)	6,13%	1,15%	0,59%
Lucro líquido sem IR (%)	93,88%	98,86%	99,41%
Lucro líquido (%)	73,88%	78,86%	79,41%
Lucro líquido (R\$)	R\$ 73,88	R\$ 788,55	R\$ 7.941,30

5.5 Resultados Experimentais

Nesta seção serão apresentados os resultados experimentais detalhados das duas melhores técnicas selecionadas para avaliação e comparação, são elas: MLP e Meta-RF. Os resultados a serem relatados dizem respeito a simulações realizadas levando-se em conta 244 dias de pregão dentro do período de Agosto de 2015 a Agosto de 2016 em 9 conjunto de dados de 9 ativos distintos da B3. Em todos os experimentos foram utilizados investimentos iniciais de R\$ 10.000,00, onde os retornos financeiros já consideram todos os custos operacionais.

A Tabela 16 apresenta os resultados simulados do melhor meta-classificador (Meta-RF), já a Tabela 17 apresenta os resultados simulados do melhor classificador isolado (MLP). Entende-se que com essas duas tabelas é possível compará-los e demonstrar características relevantes dos 2 modelos de previsão.

Ao avaliar os 9 ativos selecionados e comparar os retornos financeiros acumulados, o meta-classificador foi melhor em 7 oportunidades, perdendo somente nos ativos KROT3

e GOLL4. Com grande destaque para os ativos BOVA11, CIEL, ITUB4, USIM5, CMIG4. Os resultados são bem interessantes, visto que dentre os destaques, o meta-classificador supera o classificador isolado em mais de 100% de média, por exemplo o retorno financeiro da BOVA11 para o classificador isolado foi 4,90%, enquanto o meta-classificador foi de 27,20%. Nota-se que para o ativo ITUB4 o meta-classificador foi superior em 611%.

Para cada um dos ativos, os resultados são apresentados também nas Figuras 22, 23, 24, 25, 26, 27, 28, 29, 30; em cada Figura foram comparados as séries de retornos financeiros acumulados, os *drawdowns* e o *boxplot* dos retornos.

Nesta mesma métrica de retorno financeiro, o meta-classificador e o classificador isolado também foram comparados com 3 *Baselines*: *buy-and-hold* (BH), aleatório e estratégia inversa, conforme a Tabela 18. Ao avaliar a estratégia BH dos 9 ativos, apenas para KROT3 o meta-classificador não foi superior. Neste caso, a série de preços possui uma característica de forte tendência de alta, conforme a Figura 18. Logo, é mais frequente encontrar este comportamento em cenários de tendência de grandes altas onde a estratégia *buy-and-hold* apresenta ser superior aos modelos de previsão. Já para os demais *Baselines*, o meta-classificador foi melhor em todos os ativos.

O classificador isolado não obteve resultados satisfatórios nesta comparação; por diversas vezes foi pior que os *Baselines*, como foi o caso dos ativos: BOVA11, CIEL3, ITUB4 e PETR4.

Para as demais métricas financeiras ao comparar as Tabelas tab:resultadosMLP e 16, também é possível verificar certa superioridade do meta-classificador, como é o caso para Ganho Médio, Perda Média e *Drawdown*. Nota-se que quando o meta-classificador ganha, ganha-se mais e quando perde, perde-se menos do que o classificador isolado.

É muito importante considerar também o *Drawdown*; a Tabela 15 demonstra bem o quanto é difícil recuperar perdas. Ao simular um cenário hipotético de perdas de 50% de um capital inicial de R\$ 1000,00, o investidor teria em mãos R\$ 500,00, caso o mesmo investisse novamente este valor e tivesse uma valorização do mesmo percentual da perda, ou seja, 50%. O capital final retornaria para R\$ 750,00. Para que o capital retornasse ao valor do início da simulação, ou seja, R\$ 1000,00, haveria uma necessidade de valorização de 100%.

Neste aspecto o meta-classificador apresentou melhor performance também com um *Drawdown* médio dos ativos de -18,98% enquanto o classificador isolado obteve -22,13%.

A mesma predominância do meta-classificador também é caracterizada nas medidas de desempenho dos modelos (Acurácia, Revocação, Especificidade, F1 e Precisão).

Por fim, dentro da análise das métricas financeiras foi realizada uma comparação da dis-

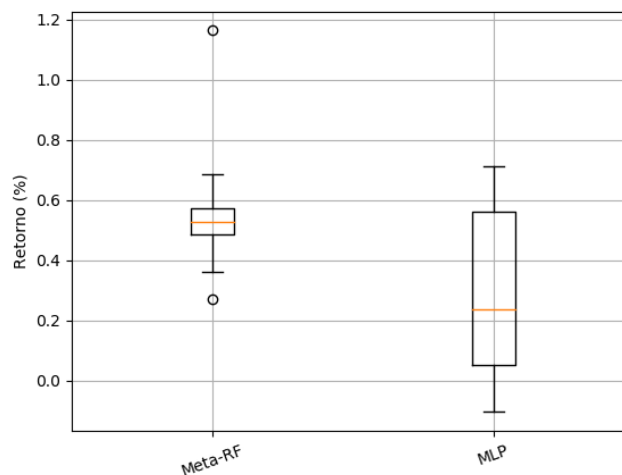
Tabela 15 – Referência de *Drawdown*.

% Capital Perdido	% Capital Requerido para Recuperar
10%	11.11%
20%	25.00%
30%	42.85%
40%	66.66%
50%	100%
60%	150%
70%	233%
80%	400%
90%	900%
100%	Falência

Fonte: Adaptado de [Zhang \(2018\)](#)

persão de todos os retornos financeiros acumulados, conforme a Figura 20, que mostra o melhor desempenho do meta-classificador. Também foi aplicado o teste t de [Student \(1908\)](#) e a partir do p-valor de 0,04015 o meta-classificador mostrou ter diferença estatística significativa, possibilitando a conclusão de que o meta-classificador é melhor que o classificador isolado.

Figura 20 – Comparação de Dispersão Geral dos Retornos Financeiros.



5.5.1 Considerações Finais

Neste capítulo foram detalhadas todas etapas dos experimentos, ou seja, como realizou-se as configurações dos algoritmos e seus parâmetros, a escolha do modelo de previsão de tendências mais adequado aos dados utilizados, a avaliação das bases de dados disponíveis e a verificação da viabilidade dos custos operacionais. Também foram apresentados os resultados do meta-classificador proposto quando comparados ao melhor classificador

Tabela 16 – Resultados Meta-Classificador - Meta-RF

Técnica	Meta-RF								
	BOVA11	CIEL3	ITUB4	PETRA	USIM5	CMIG4	GGBR4	KROT3	GOLL4
% Retorno Financeiro	27,20%	36,13%	52,62%	52,98%	116,37%	68,62%	48,51%	50,57%	57,12%
Max Drawdown	-11,54%	-10,26%	-14,82%	-22,40%	-29,59%	-18,62%	-28,80%	-9,69%	-25,06%
Volatilidade	2,52%	2,94%	7,85%	6,26%	14,18%	5,29%	6,72%	3,26%	6,59%
Índice Sharpe	6,83	8,87	5,42	6,86	7,5	11,07	5,73	12,43	7,14
% Perda Média	-1,52%	-2,08%	-2,17%	-4,14%	-5,15%	-3,21%	-4,08%	-2,33%	-3,67%
% Ganho Médio	3,41%	2,44%	4,26%	6,28%	11,90%	5,31%	6,22%	3,61%	7,97%
% Dias Positivo	8,00%	12,50%	10,00%	9,50%	9,00%	11,00%	8,50%	12,50%	7,50%
% Dias Negativo	9,00%	6,00%	7,50%	8,00%	9,50%	7,50%	7,00%	8,50%	8,50%
% Gatilhos Positivos	23,53%	33,78%	28,57%	26,39%	23,68%	29,73%	27,42%	29,76%	22,73%
% Gatilhos Negativos	26,47%	16,22%	21,43%	22,22%	25,00%	20,27%	22,58%	20,24%	25,76%
Acurácia	56,50%	53,00%	53,50%	48,00%	56,50%	54,00%	53,50%	56,00%	53,50%
Precisão	56,36%	52,88%	54,39%	51,06%	53,68%	54,26%	55,42%	56,57%	41,51%
Revocação	61,39%	55,00%	60,19%	45,28%	54,26%	51,00%	45,10%	55,45%	26,19%
Especificidade	51,52%	51,00%	46,39%	51,06%	58,49%	57,00%	62,24%	56,57%	73,28%
F1	58,77%	53,92%	57,14%	48,00%	53,97%	52,58%	49,73%	56,00%	32,12%

Tabela 17 – Resultados Classificador - MLP

Técnica	MLP								
	BOVA11	CIEL3	ITUB4	PETRA	USIM5	CMIG4	GGBR4	KROT3	GOLL4
% Retorno Financeiro	4,90%	-10,44%	8,60%	28,47%	63,37%	5,33%	23,67%	56,10%	71,13%
Max Drawdown	-10,95%	-20,70%	-12,39%	-30,03%	-14,61%	-25,77%	-29,78%	-18,50%	-36,40%
Volatilidade	2,23%	3,26%	2,58%	6,39%	4,93%	4,62%	5,98%	5,54%	14,10%
Índice Sharpe	-2,28	-6,26	-0,54	2,89	10,82	-1,008	2,28	8,32	5,75
% Perda Média	-1,50%	-2,20%	-2,28%	-4,57%	-4,03%	-3,57%	-3,50%	-2,95%	-3,71%
% Ganho Médio	2,08%	2,22%	1,71%	5,21%	5,13%	4,04%	5,60%	5,20%	12,72%
% Dias Positivo	9,50%	10,00%	14,50%	11,50%	10,50%	9,50%	9,00%	10,50%	6,50%
% Dias Negativo	11,50%	12,50%	9,00%	10,00%	5,50%	10,00%	11,00%	9,00%	10,00%
% Gatilhos Positivos	22,62%	21,74%	30,85%	26,74%	29,17%	24,36%	22,50%	26,92%	19,12%
% Gatilhos Negativos	27,38%	27,17%	19,15%	23,26%	15,28%	25,64%	27,50%	23,08%	29,41%
Acurácia	45,00%	44,50%	51,00%	54,50%	55,00%	49,00%	50,00%	56,50%	52,00%
Precisão	44,71%	44,95%	52,87%	57,73%	52,50%	49,02%	51,25%	57,61%	43,48%
Revocação	37,62%	49,00%	44,66%	52,83%	44,68%	50,00%	40,20%	52,48%	47,62%
Especificidade	52,53%	40,00%	57,73%	56,38%	64,15%	48,00%	60,20%	60,61%	55,17%
F1	40,86%	46,89%	48,42%	55,17%	48,28%	49,50%	45,05%	54,92%	45,45%

Tabela 18 – Resultado de % Retorno Financeiro dos *Baselines*

Ativo	Baseline		
	Buy-and-hold	Aléatório	Inversa
BOVA11	21,10%	5,30%	-71,40%
CIEL3	12,71%	8,52%	-80,54%
ITUB4	23,92%	14,03%	-82,16%
PETR4	52,90%	27,47%	-95,35%
USIM5	-1,63%	3,66%	-98,93%
CMIG4	38,12%	14,16%	-94,82%
GGBR4	35,23%	24,47%	-96,43%
KROT3	65,29%	33,21%	-88,64%
GOLL4	34,40%	26,56%	-98,97%

isolado e aos *Baselines* usados como referência, e em 90% dos ativos avaliados o meta-classificador mostrou melhores resultados.

Capítulo 6

Conclusão

Nesta tese foi apresentado um trabalho de pesquisa que teve por objetivo principal o desenvolvimento de um método para a previsão de séries temporais financeiras utilizando um meta-classificador que apresentasse principalmente boa performance financeira a partir de suas previsões. O objetivo proposto foi alcançado, uma vez que, de forma geral, o meta-classificador apresentou melhores resultados que os *Baselines* e os métodos concorrentes.

Todo o arcabouço projetado foi implementado na plataforma *WEKA*. Durante o desenvolvimento, diversas versões foram criadas e algumas das informações produzidas acabaram por não figurar no texto final, visto que várias destas não faziam parte diretamente do escopo final do projeto. No entanto, todas estas etapas influenciaram positivamente em diversos momentos quanto às decisões de implementação.

Inicialmente, foram investigados diversos trabalhos com o tema relacionado à previsão de séries temporais financeiras. Através de uma revisão sistemática da literatura (RSL), percebeu-se que o tema em questão é de grande interesse não só para a área acadêmica, mas também para o mercado financeiro. Não foram encontrados trabalhos que utilizam meta-classificadores com as características implementadas para este cenário proposto, o que motivou ainda mais o desenvolvimento deste trabalho.

Posteriormente, foi aplicada esta mesma técnica em dados da B3; testes iniciais foram realizados no ativo BOVA11. Porém, foram encontrados alguns problemas nas previsões, com a observação de baixa acurácia devido ao fato de ter sido utilizada apenas uma técnica, além de não existir janela deslizante. Na sequência, foi elaborada uma proposta de meta-classificação utilizando os preços absolutos como entrada com dois conjuntos de dados para treino e teste, respectivamente. Outros problemas foram detectados: um deles era os valores absolutos de preço, ou seja, a série de preços não normalizada; e o outro era que os modelos estavam apresentando características de *overfitting*. Também foi criada uma versão onde foram inseridos indicadores técnicos combinados com os preços como entrada

dos classificadores base. No entanto, não foram encontrados bons resultados em acurácia.

Na versão atual, foi desenvolvido um meta-classificador baseado em sete técnicas de inteligência computacional para descobrir tendências em séries financeiras: Programação Genética (PG), Máquinas de Vetor de Suporte (SVM), Florestas Aleatórias (RF), Redes Neurais (MLP), Árvore de Decisão (J48), Otimização Mínima Sequencial (SMO) e Redes Bayesianas (BN), o que percebemos como um grande diferencial deste trabalho.

Para validar a metodologia, realizou-se uma série de experimentos utilizando dados reais da Bolsa de Valores de São Paulo (B3). Foram utilizados 9 conjuntos de dados de ativos: BOVA11, CIEL3, ITUB4, PETR4, USIM5, CMIG4, GGBR4, KROT3 e GOLL4. Todos os dados de entrada dos algoritmos foram tratados; utilizou-se a série de retorno na escala logarítmica dos preços e volume em uma janela deslizante.

Os resultados foram satisfatórios, visto que é possível mostrar que, em geral, os meta-classificadores possibilitaram melhores ganhos, apresentando uma boa acurácia na classificação, com até 57%, além de resultados financeiros satisfatórios, com ganhos de até 100% acima do valor de capital inicialmente investido, já descontando custos operacionais.

Foi possível observar que, em geral, o modelo proposto nesse trabalho tem resultados melhores que os classificadores individuais e os *baselines*, com raras exceções. Por fim, com os resultados apresentados, pode-se concluir que os objetivos desta pesquisa foram todos alcançados e com indicativos, inclusive, de seu potencial de aplicação em cenários reais de investimentos.

6.1 Trabalhos futuros

Apesar da solução proposta ter se apresentado rentável em um ambiente simulado e com uma soluções para etapas importantes de um *algotrading* como tratamento de dados, previsão, avaliação de cenários e negociação outras não puderam ser empreendidas. Os desafios enfrentados durante as fases de testes demonstraram que certos pontos podem ser delicados neste tipo de pesquisa. Para tanto, pode-se citar, principalmente, a influência exercida por diversas variáveis como a profundidade do livro de ofertas, séries com maior granularidade, humor de mercados, relação entre horários e liquidez dos ativos, dentre outras.

Sendo assim, entende-se que o campo de estudos de previsões em séries temporais financeiras oferece uma infinidade de possibilidades de pesquisa. Desta maneira, esse trabalho pode ser estendido de diversas formas.

Pode ser que existam outros atributos que ajudem a melhorar os resultados e que podem

ser estudados, como por exemplo indicadores de sentimento de notícias relacionadas a ativos, volume de dados relacionados em redes sociais ou mesmo sobre buscas em motores de busca na *web*.

Sem muito esforço é possível a realização de testes com algoritmos de *Deep Reinforcement Learning*, com conjuntos de dados de mercado internacional e também com cripto ativos.

Do ponto de vista financeiro, é notório que a estratégia de investimento é tão importante quanto a prever corretamente o comportamento do mercado. Desta forma, entende-se que se avaliar estratégias mais robustas de negociação, possivelmente teríamos resultados melhores. Por fim, avaliar o arcabouço em um cenário real.

Referências

- ABBASI, A. et al. Metafraud: a Meta-Learning Framework for Detecting Financial Fraud. **MIS Quarterly**, v. 36, n. 4, p. 1293–1327, 2012. ISSN 02767783. Citado na página 9.
- ABILIO, R. et al. A systematic review of contemporary metrics for software maintainability. In: **Software Components Architectures and Reuse (SBCARS), 2012 Sixth Brazilian Symposium on**. [S.l.: s.n.], 2012. p. 130–139. Citado na página 78.
- ADVFN. **Entenda o que é o IGPM**. 2010. Acessado em 06 de Janeiro de 2019. Disponível em: <<https://br.advfn.com/indicadores/igpm>>. Citado na página 48.
- AGAPITOS, A.; BRABAZON, A.; O'NEILL, M. Regularised gradient boosting for financial time-series modelling. **Computational Management Science**, v. 14, n. 3, p. 367–391, 2017. Citado na página 11.
- ALMEIDA, C. **Os 10 fatos mais marcantes de 2015**. 2016. Acessado em 06 de Janeiro de 2019. Disponível em: <<https://super.abril.com.br/sociedade/os-10-fatos-mais-marcantes-de-2015/>>. Citado na página 51.
- ANGHEL, M. G. Technical analysis versus fundamental analysis of securities. **Romanian Statistical Review Supplement**, v. 61, n. 2, p. 257–262, 2013. Citado na página 16.
- ARMAKI, A. G. et al. A hybrid meta-learner technique for credit scoring of banks customers. **Engineering, Technology & Applied Science Research**, v. 7, n. 5, p. 2073–2082, 2017. Citado na página 11.
- ASHBY, W. **Design for a brain: the origin of adaptive behavior**. [S.l.]: Wiley, 1960. (Science paperbacks). Citado na página 25.
- ASSAF, A. **Mercado financeiro**. [S.l.]: Atlas, 2009. ISBN 9788522454365. Citado na página 15.
- ATSALAKIS, G. S.; VALAVANIS, K. P. Surveying stock market forecasting techniques – part ii: Soft computing methods. **Expert Systems with Applications**, v. 36, n. 3, Part 2, p. 5932 – 5941, 2009. ISSN 0957-4174. Citado na página 6.
- BANZHAF, W. et al. **Genetic Programming: An Introduction: on the Automatic Evolution of Computer Programs and Its Applications**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998. ISBN 1-55860-510-X. Citado na página 29.
- BARAK, S.; ARJMAND, A.; ORTOBELLI, S. Fusion of multiple diverse predictors in stock market. **Information Fusion**, v. 36, p. 90 – 102, 2017. ISSN 1566-2535. Citado na página 11.
- BARRYMORE, J. **Como funcionam as tendências do mercado de ações**. 2017. <<http://empresasefinancas.hsw.uol.com.br/tendencias-mercado-de-acoes.htm>>. Acessado em 06 de Fevereiro de 2017. Citado na página 1.

BASGALUPP, M. P. **Algoritmos genéticos para seleção de atributos em problemas de classificação de processos de negócio**. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio Grande do Sul, 2007. Citado na página 23.

BAYES, T. An essay towards solving a problem in the doctrine of chances. **Phil. Trans. of the Royal Soc. of London**, v. 53, p. 370–418, 1763. Citado na página 28.

BEN-HUR, A.; WESTON, J. A user's guide to support vector machines. In: _____. **Data Mining Techniques for the Life Sciences**. Totowa, NJ: Humana Press, 2010. p. 223–239. ISBN 978-1-60327-241-4. Citado na página 25.

BIONDO, A. E. et al. Are random trading strategies more successful than technical ones? **PLoS ONE**, v. 8, jul 2013. Citado na página 16.

BM&FBOVESPA. **BM&FBOVESPA**. 2017. <<http://www.bmfbovespa.com.br>>. Acessado em 06 de Fevereiro de 2017. Citado na página 15.

BORDIN, A. **Aplicação de Programação Genética na Análise de Sentimentos**. Dissertação (Mestrado) — Universidade Federal de Goiás, 2018. Citado na página 43.

BOX, G.; JENKINS, G. **Time series analysis: forecasting and control**. [S.l.]: Holden-Day, 1976. (Holden-Day series in time series analysis and digital processing). ISBN 9780816211043. Citado na página 13.

BREIMAN, L. Bagging predictors. **Mach. Learn.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 24, n. 2, p. 123–140, ago. 1996. ISSN 0885-6125. Citado 2 vezes nas páginas 11 e 35.

BREIMAN, L. Random forests. **Mach. Learn.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 45, n. 1, p. 5–32, out. 2001. ISSN 0885-6125. Citado 2 vezes nas páginas 8 e 28.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting**. 2nd. ed. [S.l.]: Springer, 2002. Hardcover. ISBN 0387953515. Citado na página 2.

BURGES, C. J. A tutorial on support vector machines for pattern recognition. **Data Mining and Knowledge Discovery**, p. 121–167, Jun 1998. Citado na página 24.

CAFFÉ, M. I. R.; PEREZ, P. S.; BARANAUSKAS, J. A. Evaluation of stacking on biomedical data. **Journal of Health Informatics**, v. 4, n. 3, 2012. Citado na página 37.

CALDEIRA, E. et al. Characterizing and evaluating fraud in electronic transactions. **Web Congress, Latin American**, IEEE Computer Society, Los Alamitos, CA, USA, v. 0, p. 115–122, 2012. Citado na página 9.

CAVALCANTE, F.; MISUMI, J. **Mercado de capitais: o que é, como funciona**. [S.l.]: Elsevier, 2005. ISBN 9788535216738. Citado na página 15.

CAVALCANTE, R. C. et al. Computational intelligence and financial markets: A survey and future directions. **Expert Systems with Applications**, v. 55, p. 194 – 211, 2016. ISSN 0957-4174. Citado 2 vezes nas páginas 1 e 6.

COHEN, W. W. Fast effective rule induction. In: **In Proceedings of the Twelfth International Conference on Machine Learning**. [S.l.]: Morgan Kaufmann, 1995. p. 115–123. Citado na página 8.

CRUZ, R. M. O. et al. META-DES: A dynamic ensemble selection framework using meta-learning. **Pattern Recognition**, v. 48, n. 5, p. 1925–1935, 2015. ISSN 00313203. Disponível em: <<http://dx.doi.org/10.1016/j.patcog.2014.12.003>>. Citado na página 8.

CUTLER, D. M.; POTERBA, J. M.; SUMMERS, L. H. **What Moves Stock Prices?** [S.l.], 1988. (Working Paper Series, 2538). Disponível em: <<http://www.nber.org/papers/w2538>>. Citado na página 16.

DABHI, V. K.; CHAUDHARY, S. Financial Time Series Modeling and Prediction Using Postfix-GP. **Computational Economics**, 2015. ISSN 0927-7099. Citado na página 8.

DAMIEN, R. **O que é o trading por algoritmos?** 2018. Acessado em 06 de Janeiro de 2019. Disponível em: <<https://www.estrategia-bolsa.pt/trading-por-algoritmos.html>>. Citado na página 19.

DAVIS, J.; GOADRICH, M. The relationship between precision-recall and roc curves. In: **Proceedings of the 23rd International Conference on Machine Learning**. New York, NY, USA: ACM, 2006. (ICML '06), p. 233–240. ISBN 1-59593-383-2. Citado na página 38.

DIETTERICH, T. G. Ensemble methods in machine learning. In: **Proceedings of the First International Workshop on Multiple Classifier Systems**. London, UK, UK: Springer-Verlag, 2000. (MCS '00), p. 1–15. ISBN 3-540-67704-6. Disponível em: <<http://dl.acm.org/citation.cfm?id=648054.743935>>. Citado na página 34.

DRUCKER, H. et al. Support vector regression machines. In: **ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 9**. [S.l.]: MIT Press, 1997. p. 155–161. Citado na página 8.

DUARTE, R. S. F.; LORENA, A. Uma aplicação de meta-aprendizagem no mercado euro/dólar. In: **WAI AF 2019 – Workshop of Artificial Intelligence Applied to Finance**. [S.l.: s.n.], 2018. p. 00–00. Citado na página 11.

DŽEROSKI, S.; ŽENKO, B. Is combining classifiers with stacking better than selecting the best one? **Machine Learning**, v. 54, n. 3, p. 255–273, 2004. ISSN 1573-0565. Citado 2 vezes nas páginas 35 e 37.

ERNST, A. Find the right people, processes and technology to manage record-to-report risks. **Managing Operational Tax Risk**, 2014. Citado na página 39.

FAMA, E. F.; MALKIEL, B. G. Efficient capital markets: A review of theory and empirical work. **The Journal of Finance**, Blackwell Publishing Ltd, v. 25, n. 2, p. 383–417, 1970. ISSN 1540-6261. Citado na página 16.

FARMER, J.; PACKARD, N. H.; PERELSON, A. S. The immune system, adaptation, and machine learning. **Physica D: Nonlinear Phenomena**, v. 22, n. 1–3, p. 187 – 204, 1986. ISSN 0167-2789. Proceedings of the Fifth Annual International Conference. Citado na página 9.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, p. 37–54, 1996. Citado na página 22.

FILHO, A.; ISHIKAWA, S. **Mercado financeiro e de capitais**. [S.l.]: Editora Atlas, 2000. ISBN 9788522424498. Citado na página 15.

FINANCIAL, E. **O que é taxa CDI e como ela funciona?** 2017. Acessado em 06 de Janeiro de 2019. Disponível em: <<https://elevenfinancial.com/o-que-e-taxa-cdi-e-como-ela-funciona>>. Citado na página 48.

FORTUNA, E. **Mercado financeiro: produtos e serviços**. [S.l.]: Qualitymark, 2007. ISBN 9788573035391. Citado na página 15.

FREITAS, A. A. **Data Mining and Knowledge Discovery with Evolutionary Algorithms**. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2002. ISBN 3540433317. Citado na página 34.

FREUND, Y. An adaptive version of the boost by majority algorithm. In: **Proceedings of the Twelfth Annual Conference on Computational Learning Theory**. New York, NY, USA: ACM, 1999. (COLT '99), p. 102–113. ISBN 1-58113-167-4. Citado na página 11.

GIACOMEL, F.; GALANTE, R.; PEREIRA, A. An algorithmic trading agent based on a neural network ensemble: A case of study in north american and brazilian stock markets. In: **2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)**. [S.l.: s.n.], 2015. v. 2, p. 230–233. Citado na página 7.

GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random forests for land cover classification. **Pattern Recognition Letters**, v. 27, n. 4, p. 294 – 300, 2006. ISSN 0167-8655. Citado na página 28.

GLANTZ, M.; KISSELL, R. **Multi-asset Risk Modeling: Techniques for a global economy in an electronic and algorithmic trading era**. 1. ed. [S.l.]: Academic Press, 2013. ISBN 978-0-12-401690-3. Citado na página 16.

GOLDBERG, D. E. **Genetic Algorithms in Search, Optimization and Machine Learning**. New York, NY, USA: Addison-Wesley Longman Publishing Co., Inc., 1989. Citado 3 vezes nas páginas 29, 32 e 33.

GOLDSCHMIDT, R. R. **Uma Introdução à Inteligência Computacional: fundamentos, ferramentas e aplicações**. Rio de Janeiro, RJ, BR: IST-Rio, 2010. Citado 3 vezes nas páginas 2, 21 e 22.

GOOGLE. **Google**. 2017. <<https://www.google.com.br/>>. Acessado em 31 de Janeiro de 2017. Citado na página 14.

GRAF BOLSA. **Indicadores Técnicos**. 2017. <<http://www.grafbolsa.com/help/indtec.html>>. Acessado em 06 de Fevereiro de 2017. Citado na página 18.

GRAHAM, B. Some calculus suggestions by a student. **The American Mathematical Monthly**, Mathematical Association of America, v. 24, n. 6, p. 265–271, 1917. ISSN 00029890, 19300972. Disponível em: <<http://www.jstor.org/stable/2973181>>. Citado na página 6.

HAMILTON, W.; DOW, C. **The Stock Market Barometer**. [S.l.]: Harper & Brothers, 1922. Citado na página 6.

HAND, D. J.; YU, K. Idiot's bayes—not so stupid after all? **International Statistical Review**, Blackwell Publishing Ltd, v. 69, n. 3, p. 385–398, 2001. ISSN 1751-5823. Disponível em: <<http://dx.doi.org/10.1111/j.1751-5823.2001.tb00465.x>>. Citado na página 37.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. 2. ed. New York, NY, USA: Springer, 2009. Disponível em: <<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>>. Citado na página 3.

HEBB, D. O. **The Organization of Behavior: A Neuropsychological Theory**. New ed. New York: Wiley, 1949. Hardcover. ISBN 0805843000. Citado na página 25.

HO, T. K. Multiple classifier combination: Lessons and next steps. In: _____. [S.l.]: World Scientific, 2001. (Series in Machine Perception and Artificial Intelligence, v. 47), cap. 7, p. 171–198. Citado na página 35.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Comput.**, MIT Press, Cambridge, MA, USA, v. 9, n. 8, p. 1735–1780, nov. 1997. ISSN 0899-7667. Citado na página 7.

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems**. Ann Arbor, MI, USA: University of Michigan Press, 1975. Citado na página 29.

INFOMONEY. **Ações: análise técnica ajuda a determinar o momento para entrar ou sair do mercado**. 2017. <<http://www.infomoney.com.br/educacao/guias/noticia/360077/acoes-analise-tecnica-ajuda-determinar-momento-para-entrar-sair-mercado/>>. Acessado em 06 de Fevereiro de 2017. Citado na página 1.

INVESTIDOR, B. do. **Bússola do Investidor**. 2017. <<https://www.bussoladoinvestidor.com.br/>>. Acessado em 31 de Janeiro de 2017. Citado 4 vezes nas páginas 17, 19, 20 e 46.

JOHNSTON, S. A. **Trading Options to Win: Profitable Strategies and Tactics for Any Trader**. [S.l.]: Wiley, 2003. ISBN 0471226858. Citado na página 4.

JONG, K. A. D. **An Analysis of the Behavior of a Class of Genetic Adaptive Systems**. Tese (Doutorado) — University of Michigan, Ann Arbor, MI, USA, 1975. AAI7609381. Citado na página 33.

KAMPOURIDIS, M.; OTERO, F. E. B. Heuristic procedures for improving the predictability of a genetic programming financial forecasting algorithm. **Soft Computing**, 2015. ISSN 14337479. Disponível em: <<http://link.springer.com/10.1007/s00500-015-1614-8>>. Citado na página 7.

KIRKPATRICK, C.; DAHLQUIST, J. **Technical Analysis: The Complete Resource for Financial Market Technicians**. First. [S.l.]: FT Press, 2006. ISBN 0131531131. Citado na página 17.

KITCHENHAM, B. **Procedures for performing systematic reviews**. [S.l.], 2004. Citado 3 vezes nas páginas 6, 77 e 80.

KOZA, J. R. **Genetic Programming: On the Programming of Computers by Means of Natural Selection**. Cambridge, MA, USA: MIT Press, 1992. ISBN 0-262-11170-5. Citado 7 vezes nas páginas 29, 30, 31, 32, 33, 34 e 35.

KUNCHEVA, L. I. **Combining Pattern Classifiers: Methods and Algorithms**. New York, NY, USA: Wiley-Interscience, 2004. ISBN 0471210781. Citado 3 vezes nas páginas 34, 35 e 36.

- LECUN, Y. et al. Gradient-based learning applied to document recognition. In: **Proceedings of the IEEE**. [S.l.: s.n.], 1998. p. 2278–2324. Citado na página 7.
- LICHMAN, M. **UCI Machine Learning Repository**. 2013. Disponível em: <<http://archive.ics.uci.edu/ml>>. Citado na página 8.
- LIMA, L. **A difícil senão impossível tarefa de prever o futuro**. 2017. <<http://carodinho.blogfolha.uol.com.br/2013/04/11/a-dificil-senao-impossivel-tarefa-de-prever-o-futuro/>>. Acessado em 06 de Fevereiro de 2017. Citado na página 2.
- LO. **Stacking, Blending and Stacked Generalization**. 2017. <<http://www.chioka.in/stacking-blending-and-stacked-generalization/>>. Acessado em 06 de Fevereiro de 2017. Citado na página 37.
- LO, A.; MACKINLAY, A. **A non-random walk down Wall Street**. Princeton, NJ [u.a.]: Princeton Univ. Press, 1999. ISBN 0691057745. Citado na página 16.
- LUKE, S. et al. **ECJ 16: A Java-based Evolutionary Computation Research System**. <http://cs.gmu.edu/~eclab/projects/ecj/>: [s.n.], 2007. Citado na página 8.
- MAKRIDAKIS, S.; HIBON, M. The m3-competition: results, conclusions and implications. **International Journal of Forecasting**, v. 16, n. 4, p. 451 – 476, 2000. ISSN 0169-2070. The M3 - Competition. Citado na página 10.
- MALKIEL, B. G. **A Random Walk Down Wall Street**. [S.l.]: Norton, New York, 1973. Citado na página 16.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115–133, 1943. ISSN 1522-9602. Citado na página 25.
- METATRADER. **Plataforma de negociação MetaTrader 5**. 2017. <<https://www.metatrader5.com/>>. Acessado em 06 de Fevereiro de 2017. Citado na página 42.
- MISHRA, A.; POKHREL, P.; HOQUE, M. T. Stackdppred: A stacking based prediction of dna-binding protein from sequence. 2018. Citado na página 10.
- MITCHELL, T. **Machine Learning**. [S.l.]: McGraw-Hill, 1997. (McGraw-Hill International Editions). ISBN 9780071154673. Citado na página 27.
- MONTANA, D. J. Strongly typed genetic programming. **Evolutionary Computation**, v. 3, p. 199–230, 1994. Citado na página 31.
- MYKOVA, R.; HAJEK, P.; OLEJ, V. Predicting abnormal stock return volatility using textual analysis of news a meta-learning approach. **Amfiteatru Economic**, v. 20, n. 47, p. 185–201, 2018. ISSN 1582-9146. Citado na página 10.
- NAMETALA, C. A. et al. An automated investment strategy using artificial neural networks and econometric predictors. In: . Porto Alegre, Brazil: Brazilian Computer Society, 2016. (SBSI 2016), p. 21:152–21:159. ISBN 978-85-7669-317-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=3021955.3021982>>. Citado na página 7.
- NEAPOLITAN, R. E. **Learning Bayesian Networks**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2003. ISBN 0130125342. Citado na página 29.

NORONHA, M. **Análise Técnica: Teorias Ferramentas Estratégias**. [S.l.]: Editec, 2003. Citado na página 17.

OLIVEIRA, M. **Corretoras com menor taxa de corretagem**. 2019. Disponível em: <<https://blog.magnetis.com.br/corretoras-com-menor-taxa-de-corretagem/>>. Citado na página 59.

PATEL, J. et al. Predicting stock market index using fusion of machine learning techniques. **Expert Systems with Applications**, v. 42, n. 4, p. 2162–2172, 2015. ISSN 09574174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417414006551>>. Citado na página 8.

PEARL, J. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988. ISBN 0-934613-73-7. Citado na página 28.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **J. Mach. Learn. Res.**, JMLR.org, v. 12, p. 2825–2830, nov. 2011. ISSN 1532-4435. Disponível em: <<http://dl.acm.org/citation.cfm?id=1953048.2078195>>. Citado na página 26.

PEREIRA, M. d. A. **Classificação de Dados Híbridos Através de Algoritmos Evolucionários**. Tese (Doutorado) — Universidade Federal de Minas Gerais, 2012. Citado 3 vezes nas páginas 32, 33 e 38.

PERSIO, L. D.; HONCHAR, O. Artificial neural networks approach to the forecast of stock market price movements. **International Journal of Economics and Management Systems**, v. 1, p. 158–162, 2016. Citado na página 7.

PIMENTA, A. et al. GoldMiner: A genetic programming based algorithm applied to Brazilian Stock Market. **2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)**, p. 397–402, 2014. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7008695>>. Citado 2 vezes nas páginas 8 e 17.

PLATT, J. Fast training of support vector machines using sequential minimal optimization. In: **Advances in Kernel Methods - Support Vector Learning**. [S.l.]: MIT Press, 1998. Citado na página 27.

PLATT, J. C. Fast training of support vector machines using sequential minimal optimization. In: **Advances in Kernel Methods**. Cambridge, MA, USA: MIT Press, 1999. p. 185–208. ISBN 0-262-19416-3. Disponível em: <<http://dl.acm.org/citation.cfm?id=299094.299105>>. Citado na página 24.

QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 1-55860-238-0. Citado 2 vezes nas páginas 8 e 26.

RAMOS, J. P. S. Fruit sorting using artificial neural networks: bidimensional case. **Ciência e Agrotecnologia**, scielo, v. 27, p. 356 – 362, 04 2003. ISSN 1413-7054. Citado na página 26.

READHEAD, J. **What is a Support Vector Machine?** 2017. <<https://www.mql5.com/en/articles/584>>. Acessado em 06 de Fevereiro de 2017. Citado na página 23.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, p. 65–386, 1958. Citado na página 25.

SANTOS, E. M. d. **Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência**. Dissertação (Mestrado) — Universidade Federal da Paraíba, 2002. Citado na página 22.

SCHAPIRE, R. E. The strength of weak learnability. **Mach. Learn.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 5, n. 2, p. 197–227, jul. 1990. ISSN 0885-6125. Disponível em: <<http://dx.doi.org/10.1023/A:1022648800760>>. Citado 2 vezes nas páginas 11 e 35.

SEEWALD, A. K. How to make stacking better and faster while also taking care of an unknown weakness. In: **Proceedings of the Nineteenth International Conference on Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2002. (ICML '02), p. 554–561. ISBN 1-55860-873-7. Disponível em: <<http://dl.acm.org/citation.cfm?id=645531.656165>>. Citado na página 10.

SEKER, S. E. et al. Ensemble classification over stock market time series and economy news. In: **IEEE ISI 2013 - 2013 IEEE International Conference on Intelligence and Security Informatics: Big Data, Emergent Threats, and Decision-Making in Security Informatics**. [S.l.: s.n.], 2013. p. 272–273. ISBN 9781467362115. Citado na página 9.

SOLTANI, N.; AKBARI, M.; JAVAN, M. A new user-based model for credit card fraud detection based on artificial immune system. In: **6th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)**. [S.l.: s.n.], 2012. p. 029–033. Citado na página 9.

STROUSTRUP, B. **The C++ Programming Language**. 3rd. ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2000. ISBN 0201700735. Citado na página 42.

STRUZIK, Z. R. Wavelet methods in (financial) time-series processing. **Physica A: Statistical Mechanics and its Applications**, v. 296, n. 1–2, p. 307 – 319, 2001. ISSN 0378-4371. Citado na página 7.

STUDENT. The probable error of a mean. **Biometrika**, v. 6, n. 1, p. 1–25, 03 1908. ISSN 0006-3444. Citado na página 61.

TALAGALA, T. S.; HYNDMAN, R. J.; ATHANASOPOULOS, G. **Meta-learning how to forecast time series**. [S.l.], 2018. Citado na página 10.

TAN, A. C.; GILBERT, D. Ensemble machine learning on gene expression data for cancer classification. University of Glasgow, 2003. Citado na página 35.

TKáč, M.; VERNER, R. Artificial neural networks in business: Two decades of research. **Applied Soft Computing**, v. 38, p. 788 – 804, 2016. ISSN 1568-4946. Citado 2 vezes nas páginas 1 e 6.

TSAY, R. S. **Analysis of financial time series**. Hoboken (N.J.): Wiley-Interscience, 2005. (Wiley series in probability and statistics). ISBN 0-471-69074-0. Disponível em: <<http://opac.inria.fr/record=b1127642>>. Citado 2 vezes nas páginas 13 e 14.

VAPNIK, V. N. **The Nature of Statistical Learning Theory**. New York, NY, USA: Springer-Verlag New York, Inc., 1995. ISBN 0-387-94559-8. Citado na página 22.

VAPNIK, V. N.; CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. **Theory of Probability and its Applications**, SIAM, v. 16, n. 2, p. 264–280, 1971. Citado na página 22.

VARGA, G. Índice de sharpe e outros indicadores de performance aplicados a fundos de ações brasileiros. **Revista de Administração Contemporânea**, 2001. Citado na página 39.

VIEIRA, D. **Custos operacionais – saiba quais são as taxas cobradas no mercado de ações**. 2015. Acessado em 06 de Janeiro de 2019. Disponível em: <<http://daltonvieira.com/custos-operacionais-saiba-quais-sao-as-taxas-cobradas-no-mercado-de-acoes>>. Citado na página 20.

WERBOS, P. J. **The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting**. New York, NY, USA: Wiley-Interscience, 1994. ISBN 0-471-59897-6. Citado na página 25.

WIDROW, B.; HOFF, M. E. Adaptive switching circuits. In: **1960 IRE WESCON Convention Record, Part 4**. New York: IRE, 1960. p. 96–104. Citado 2 vezes nas páginas 7 e 25.

WIDROW, B.; HOFF, M. E. Associative storage and retrieval of digital information in networks of adaptive “neurons”. In: _____. Boston, MA: Springer US, 1962. p. 160–160. ISBN 978-1-4684-1716-6. Citado na página 25.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005. ISBN 0120884070. Citado na página 37.

WITTEN, I. H. et al. **Weka: Practical Machine Learning Tools and Techniques with Java Implementations**. 1999. Citado na página 22.

WOLPERT, D. H. Stacked generalization. **Neural Networks**, Pergamon Press, v. 5, p. 241–259, 1992. Citado 3 vezes nas páginas 3, 35 e 37.

YERIMA, S. Y.; SEZER, S. Droidfusion: A novel multilevel classifier fusion approach for android malware detection. **IEEE Transactions on Cybernetics**, p. 1–14, 2018. ISSN 2168-2267. Citado na página 10.

Zenko, B.; Todorovski, L.; Dzeroski, S. A comparison of stacking with meta decision trees to bagging, boosting, and stacking with other methods. In: **Proceedings 2001 IEEE International Conference on Data Mining**. [S.l.: s.n.], 2001. p. 669–670. Citado na página 37.

ZHANG, H. **Stochastic Drawdowns**. [S.l.]: World Scientific Publishing Co. Pte. Ltd., 2018. Citado na página 61.

ZHANG, Z. Z. M. A review on multi-label learning algorithms. **IEEE Transactions on Knowledge and Data Engineering**, 2014. Citado na página 37.

Apêndices

APÊNDICE A – Revisão Sistemática da Literatura (RSL)

[Kitchenham \(2004\)](#) define a Revisão Sistemática da Literatura (RSL) como um meio de identificar, avaliar e interpretar pesquisas disponíveis relevantes para a questão de pesquisa específica, área temática ou fenômeno de interesse.

As principais vantagens da RSL são:

- Possuir uma metodologia bem definida e menos tendenciosa que fornece informações sobre os efeitos de algum fenômeno dada uma variedade de configurações e métodos empíricos;
- Possibilitar consolidar as evidências e resultados obtidos em estudos anteriores sobre o tema de interesse;
- Permitir identificar lacunas na teoria e pesquisas recentes como embasamento para o aprimoramento das pesquisas;
- Fornecer embasamento e modelos teóricos para posicionar apropriadamente novos temas e oportunidades de pesquisa, refutar ou validar hipóteses, ou criar novas hipóteses sobre um determinado tema de pesquisa.

Uma RSL possui um eixo com três etapas principais em sua estrutura. A Figura 21 mostra essas etapas.

- **Planejamento:** Nesta fase é desenvolvido o objetivo da revisão da literatura e o protocolo da revisão sistemática;
- **Execução:** É conduzido o processo de revisão da literatura com a seleção dos estudos primários, verificação de sua qualidade, coleta de dados e síntese dos resultados;
- **Resultados:** Formatação e avaliação do relatório principal da revisão sistemática.

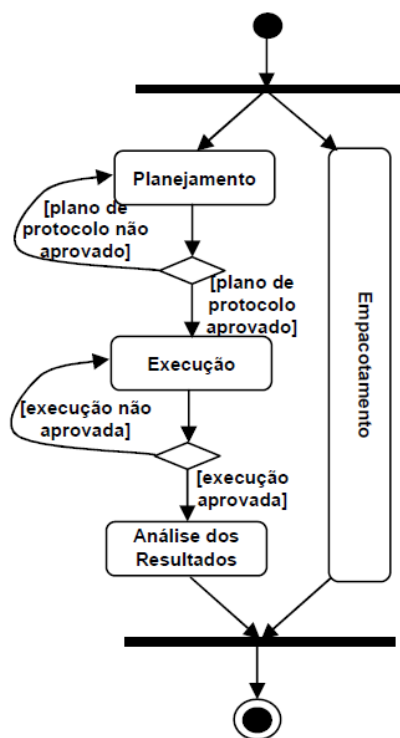
A seguir, será detalhada cada etapa do processo de revisão sistemática da literatura aplicado ao trabalho em questão.

A.0.1 Planejamento

Para [Kitchenham \(2004\)](#) o planejamento é o momento de definir o tema da pesquisa, suas questões e o protocolo de avaliação.

O tema principal desse trabalho é o desenvolvimento de um meta-classificador baseado em

Figura 21 – Principais Etapas da Revisão Sistemática da Literatura.



Fonte: [Abilio et al. \(2012\)](#)

inteligência computacional para predição de séries temporais financeiras.

Quanto às questões, definiu-se como objetivo responder às seguintes questões de pesquisa (QP) decompostas do tema principal:

- **QP1:** Quais são as bases e ou fontes de dados disponíveis e como realizar a extração / coleta destes dados?
- **QP2:** Quais são as estratégias / algoritmos de negociação existentes para o mercado de ações?
- **QP3:** Como tratar o dado bruto, a aplicação de normalização, técnicas de agrupamento e a redução de dimensionalidade?
- **QP4:** Quais são as técnicas e ou modelos mais adequados, suas configurações e parâmetros?
- **QP5:** Quais são as medidas de desempenho e de avaliação de modelos de predição?
- **QP6:** Quais são os trabalhos existentes que utilizam técnicas para predição de séries temporais financeiras?
- **QP7:** Quais são os trabalhos existentes que utilizam técnicas de inteligência computacional para predição de séries temporais financeiras?
- **QP8:** Quais são os trabalhos existentes que utilizam meta-classificadores?
- **QP9:** Quais são os trabalhos existentes que utilizam meta-classificadores para predi-

ção de séries temporais financeiras?

Assim, para apoiar a construção e direcionamento das questões de pesquisa, é necessário estruturar um protocolo de avaliação. Este protocolo descreve o processo, técnicas e ferramentas que serão utilizadas durante a fase 2 (Execução), que, por fim, irá gerar as saídas.

O protocolo é composto dos seguintes critérios:

- **Base de dados (BD) e ferramentas de busca (FB):** possíveis bases de dados candidatas que possam fornecer trabalhos relacionados em relação ao projeto proposto;
- **Critérios de busca (CB):** possíveis *strings* de busca (SB) que fazem parte do contexto do trabalho;
- **Critérios de inclusão (CI):** é a definição de quais tipos de trabalhos (artigos, dissertações, livros e periódicos) serão coletados;
- **Critérios de exclusão (CE):** consiste em uma análise qualitativa dos trabalhos coletados, onde é verificada a disponibilidade dos textos, título, tema e resumo coerentes com o contexto proposto do trabalho, com a flexibilidade de diminuir ou aumentar a granularidade da investigação.
- **Critérios de qualidade (CQ):** é o refinamento de todos os trabalhos coletados com o objetivo de selecionar por categoria os trabalhos mais relevantes que se identificam com o propósito do tema da pesquisa;
- **Dados coletados (DC):** após a triagem da etapa anterior, as informações dos trabalhos coletados são catalogadas para posterior consulta em relação aos parâmetros de publicação, como, por exemplo, ano, tipo, título, autores, etc., com o foco de categorizar os trabalhos selecionados.

Segue o protocolo desenvolvido neste trabalho.

- **Base de dados e ferramentas de busca:**
 - **BD1:** *IEEE conference and journals*;
 - **BD2:** *Google scholar*;
 - **BD3:** Google;
 - **BD4:** *ACM conference and journals*;
 - **BD5:** *Springer*;
 - **BD5:** *Elsevier*;
 - **BD6:** Banco de teses e dissertações da Capes;
 - **BD7:** Repositório dos grupos de pesquisas internos (CEFET-MG e UFMG);
 - **BD8:** *Scopus*;
- **Critérios de busca:**
 - **SB1:** (genetic programming **OR** genetic algorithm **OR** computational intelligence)

AND (stock market or ensemble **OR** classifiers **OR** meta-classifier **OR** timeseries **OR** forecasting applications **OR** stock exchange or financial market **OR** financial indicators)

- **SB2**: (programação genética **OR** algoritmo genético **OR** inteligência computacional) **AND** (mercado de ações **OR** meta-classificador **OR** classificadores **OR** series temporais **OR** previsão **OR** bolsa de valores **OR** mercado financeiro **OR** indicadores financeiros);

- **Critérios de inclusão:**

- **CI1**: artigos publicados em revistas científicas;
- **CI2**: artigos publicados em anais de congressos;
- **CI3**: dissertações defendidas;
- **CI4**: teses defendidas;
- **CI5**: relatórios de pesquisas de entidades idôneas;
- **CI6**: capítulos de livros;

- **Critérios de exclusão:**

- **CE1**: trabalhos que não disponibilizam o texto completo para leitura;
- **CE2**: trabalhos repetidos;
- **CE3**: trabalhos sem resumo;

- **Critérios de qualidade:**

- **CQ1**: objetivos geral e específicos estão explícitos?
- **CQ2**: a metodologia do trabalho está clara?
- **CQ3**: existem resultados consistentes?

- **Dados coletados:**

- **DC1**: autores
- **DC3**: título
- **DC4**: resumo
- **DC5**: ano de publicação
- **DC6**: tipo de publicação
- **DC7**: palavras-chave

A.0.2 Execução

Após a etapa de planejamento, inicia-se o processo de seleção e filtros dos trabalhos encontrados a partir dos critérios estabelecidos no protocolo.

Segundo [Kitchenham \(2004\)](#) esta etapa de execução compreende cinco estágios:

- **Identificação da pesquisa**: os critérios de busca são testados e refinados até que se tenha um conjunto de publicações satisfatório para a pesquisa.
- **Seleção dos estudos primários**: é feita por meio da leitura dos títulos e resumos das

publicações apresentadas como resultado da busca do estágio anterior. Os trabalhos selecionados são arquivados no software de gestão de referências, Mendeley¹.

- **Avaliação da qualidade dos estudos coletados:** a avaliação das publicações selecionadas no estágio anterior é realizada aplicando os critérios de qualidade.
- **Extração e monitoração de dados:** neste estágio, as publicações são inteiramente lidas e os dados são coletados.
- **Síntese dos dados:** as publicações são classificadas, de acordo com os dados coletados.

A.0.3 Análise dos Resultados

O último momento da revisão é uma compilação das informações coletadas em um texto final no qual são apresentados os resultados encontrados, com as devidas análises conforme o objetivo da pesquisa.

Como resultado da revisão sistemática de literatura, foram selecionadas as publicações conforme a metodologia proposta. A Tabela 19 apresenta o resultado final.

Tabela 19 – Resultado RSL: Quantidade de trabalhos selecionados por área pesquisada.

Técnica	Quantidade
Inteligência Computacional em Séries Temporais Financeiras	342
Meta-Classificadores	42
Meta-Classificador em Séries Temporais Financeiras	4

Durante toda a pesquisa de RSL foram encontrados diversos trabalhos publicados sobre previsão de séries temporais financeiras, isso mostra que o campo de pesquisa ainda é muito forte, visto a grande dificuldade de prever tais tendências. Foram encontrados poucos trabalhos que aplicassem meta-classificadores para Previsão de Séries Temporais Financeiras.

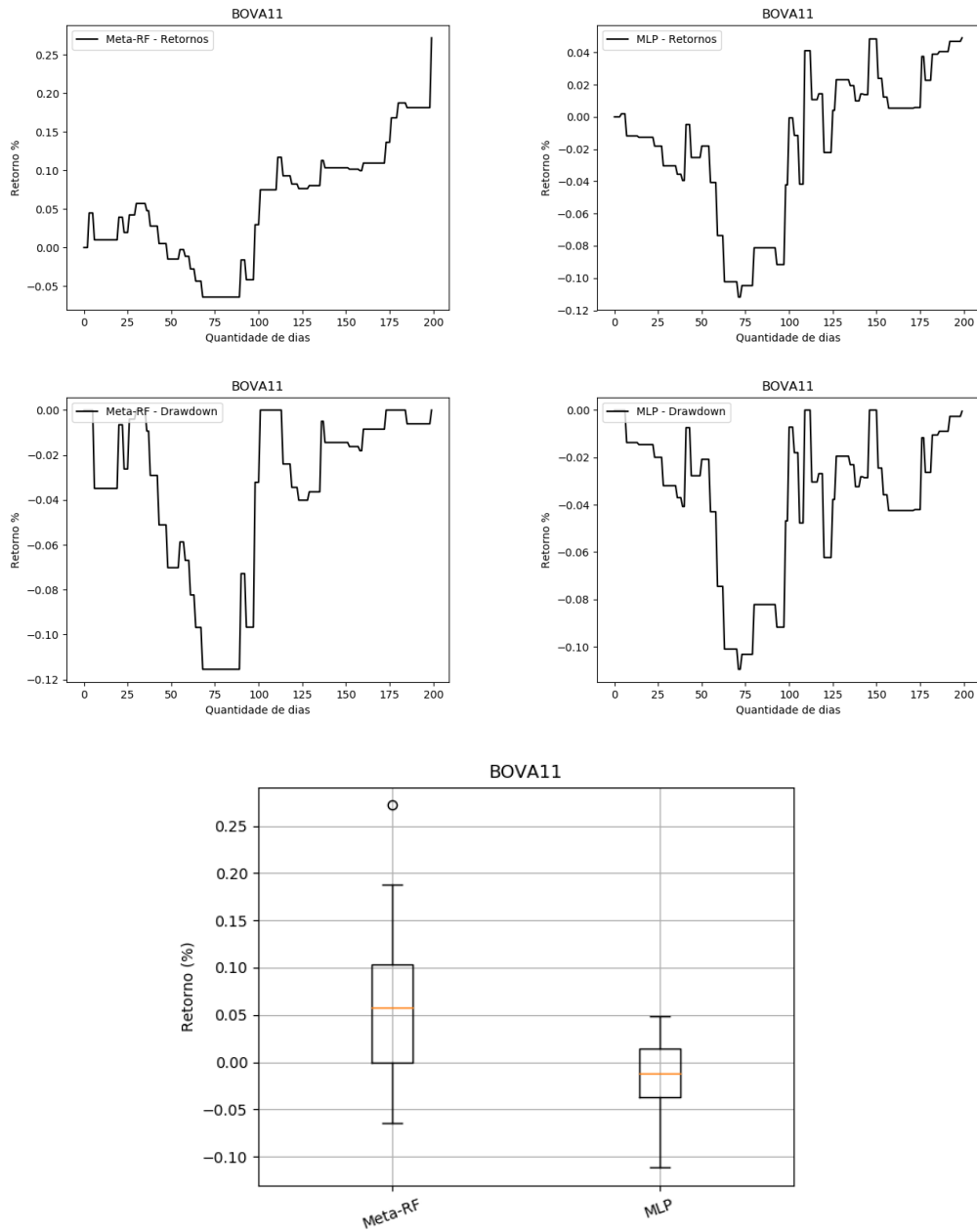
Assim, é possível interpretar tal cenário como uma grande oportunidade de aplicação deste tema em um trabalho de doutorado. Tal trabalho teria como objetivo o desenvolvimento de novos caminhos, estudos e pesquisas inovadoras no âmbito do mercado financeiro.

Por fim, a RSL contribuiu de forma sistemática e metódica no intuito de responder todas as questões de pesquisa (QP1, QP2, QP3, QP4, QP5, QP6, QP7, QP8 e QP9).

¹<http://www.mendeley.com/>

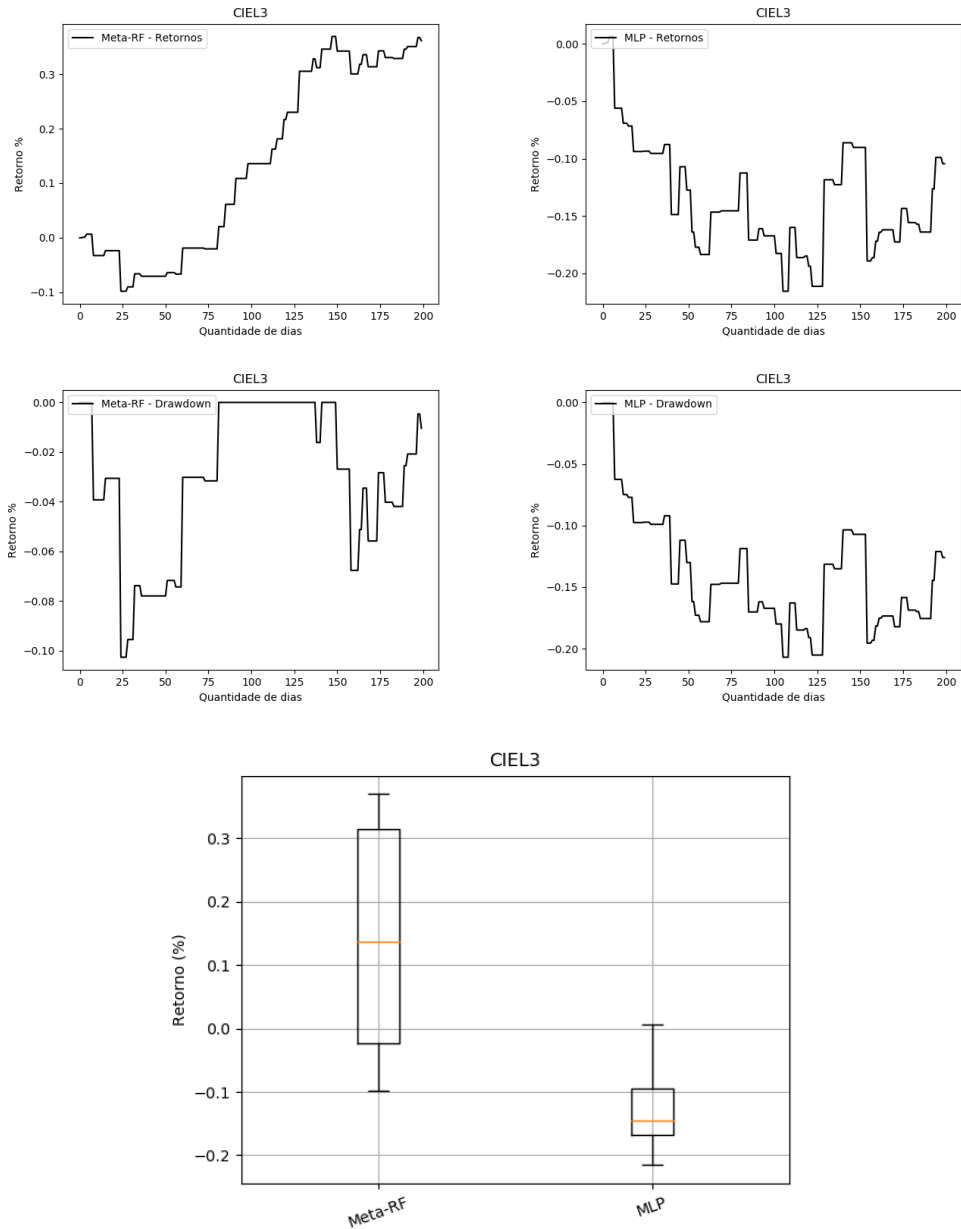
APÊNDICE B – BOVA11

Figura 22 – Comparação Drawdown e % Retornos Financeiros - BOVA11.



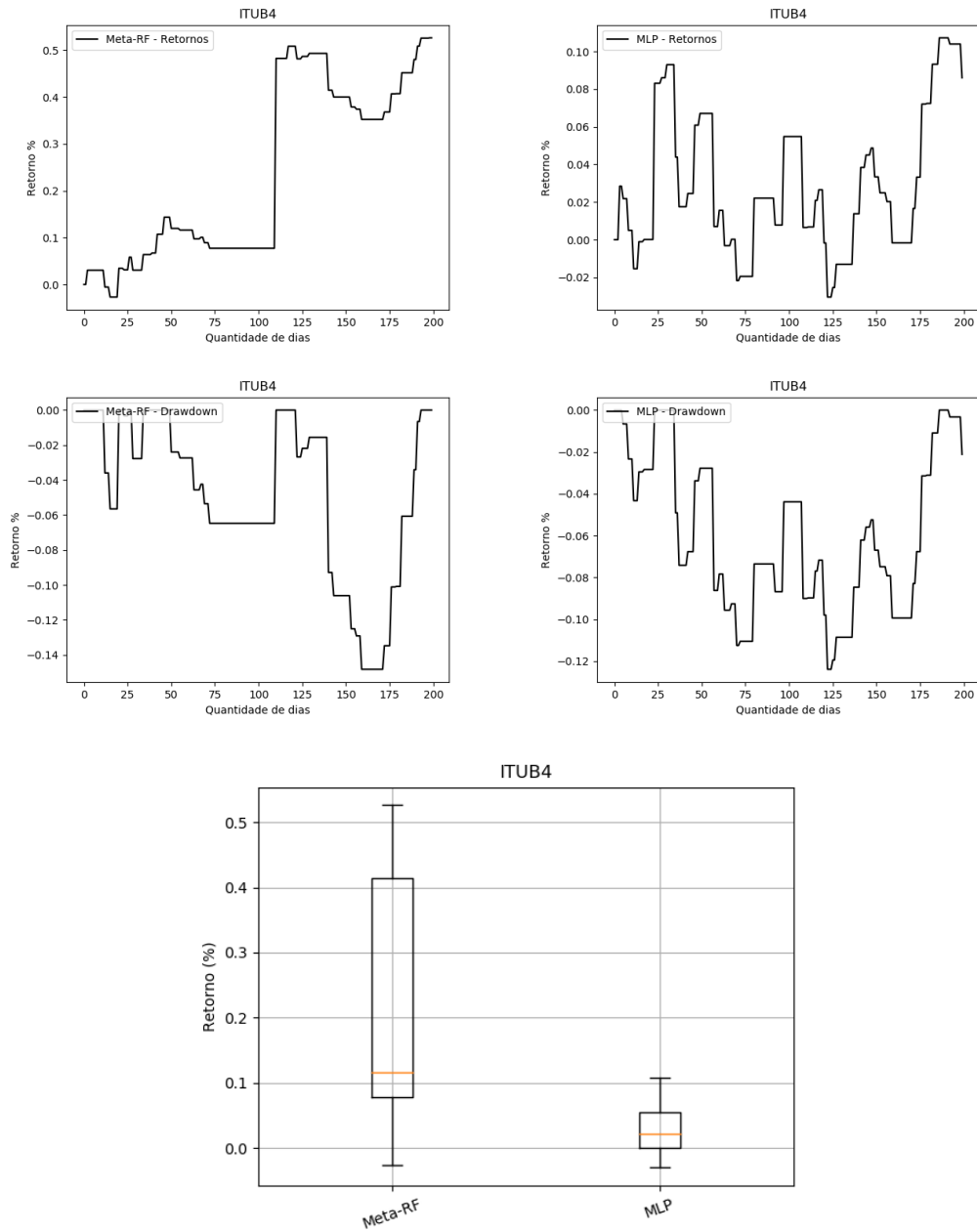
APÊNDICE C – CIEL3

Figura 23 – Comparação Drawdown e % Retornos Financeiros - CIEL3.



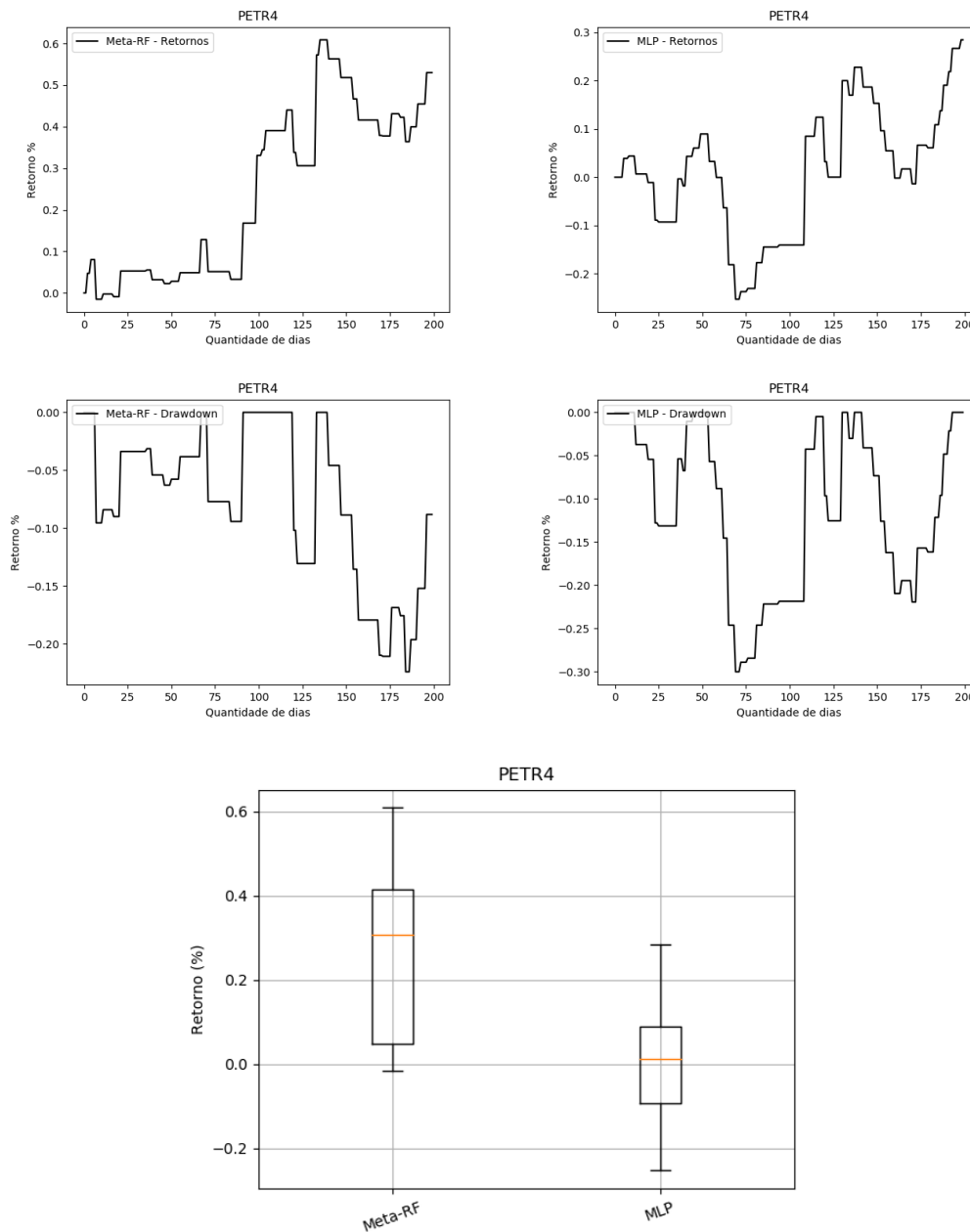
APÊNDICE D – ITUB4

Figura 24 – Comparação Drawdown e % Retornos Financeiros - ITUB4.



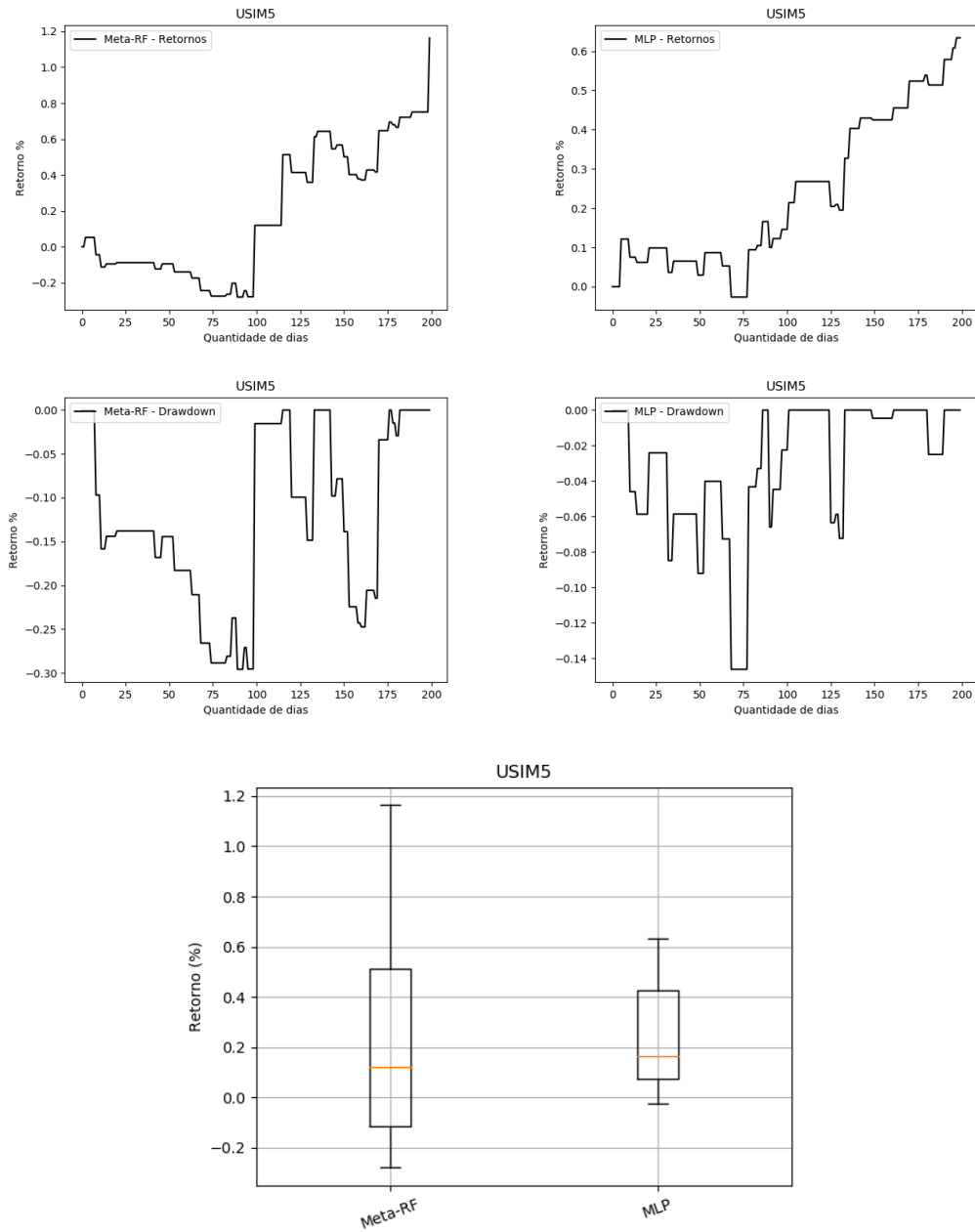
APÊNDICE E – PETR4

Figura 25 – Comparação Drawdown e % Retornos Financeiros - PETR4.



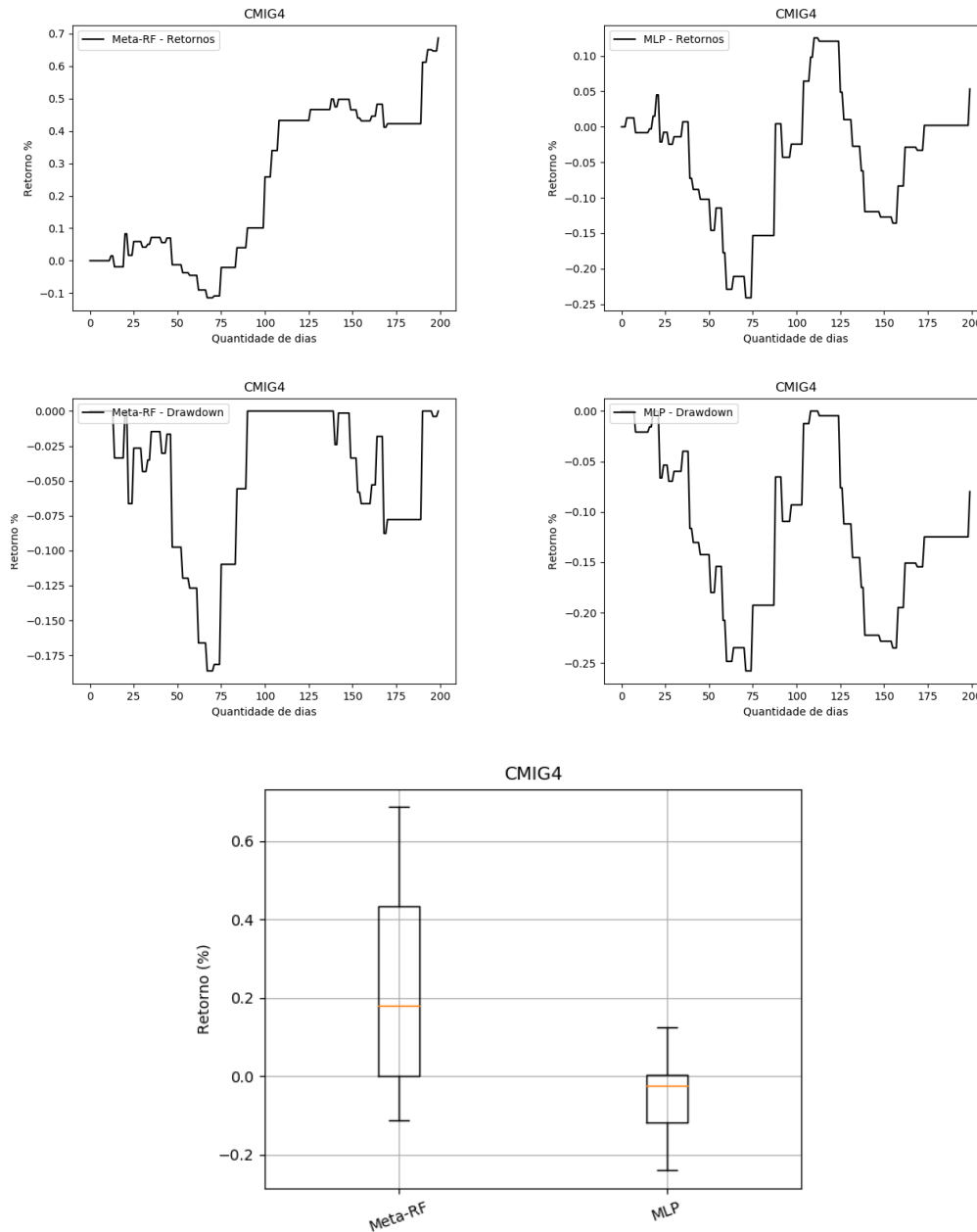
APÊNDICE F – USIM5

Figura 26 – Comparação Drawdown e % Retornos Financeiros - USIM5.



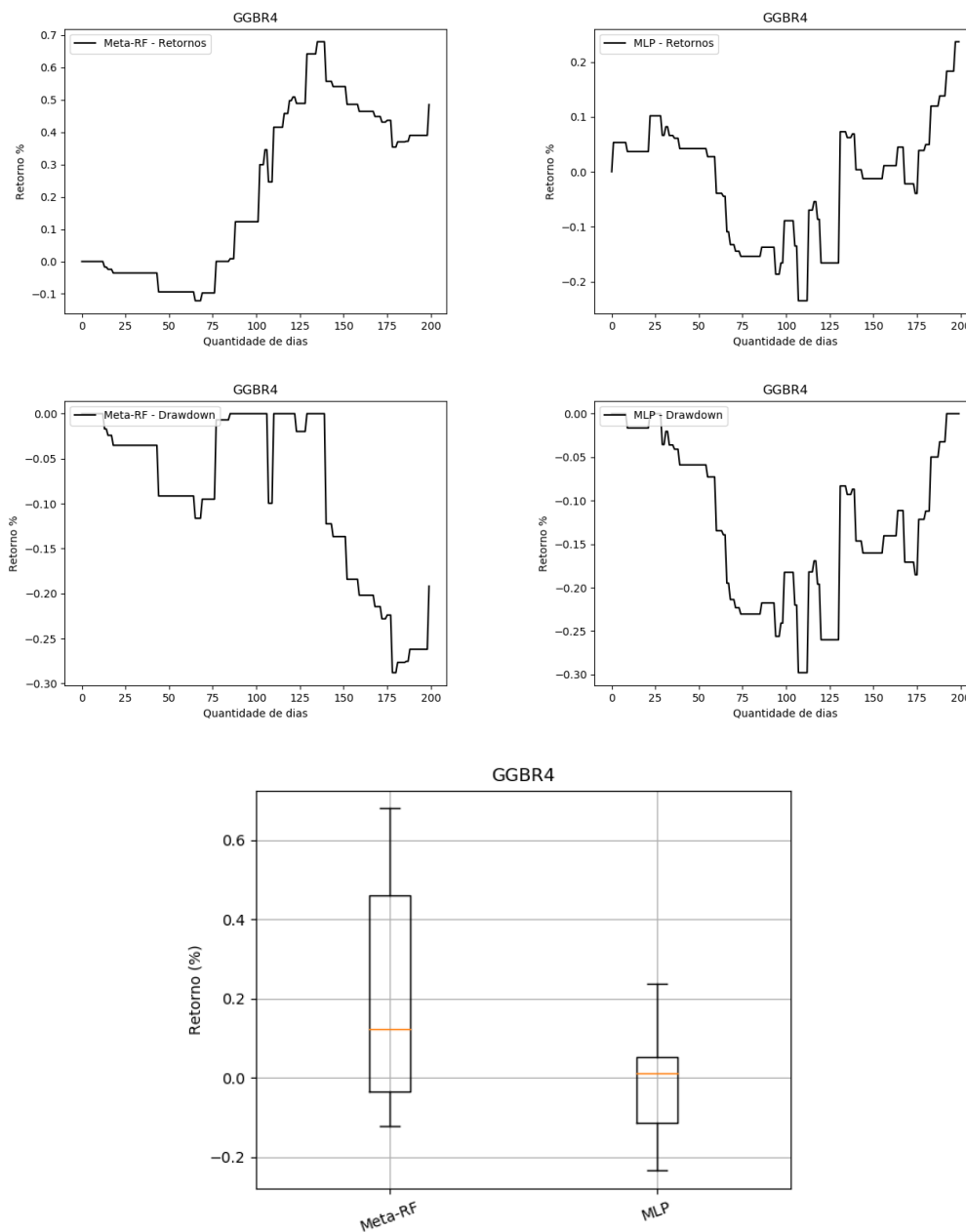
APÊNDICE G – CMIG4

Figura 27 – Comparação Drawdown e % Retornos Financeiros - CMIG4.



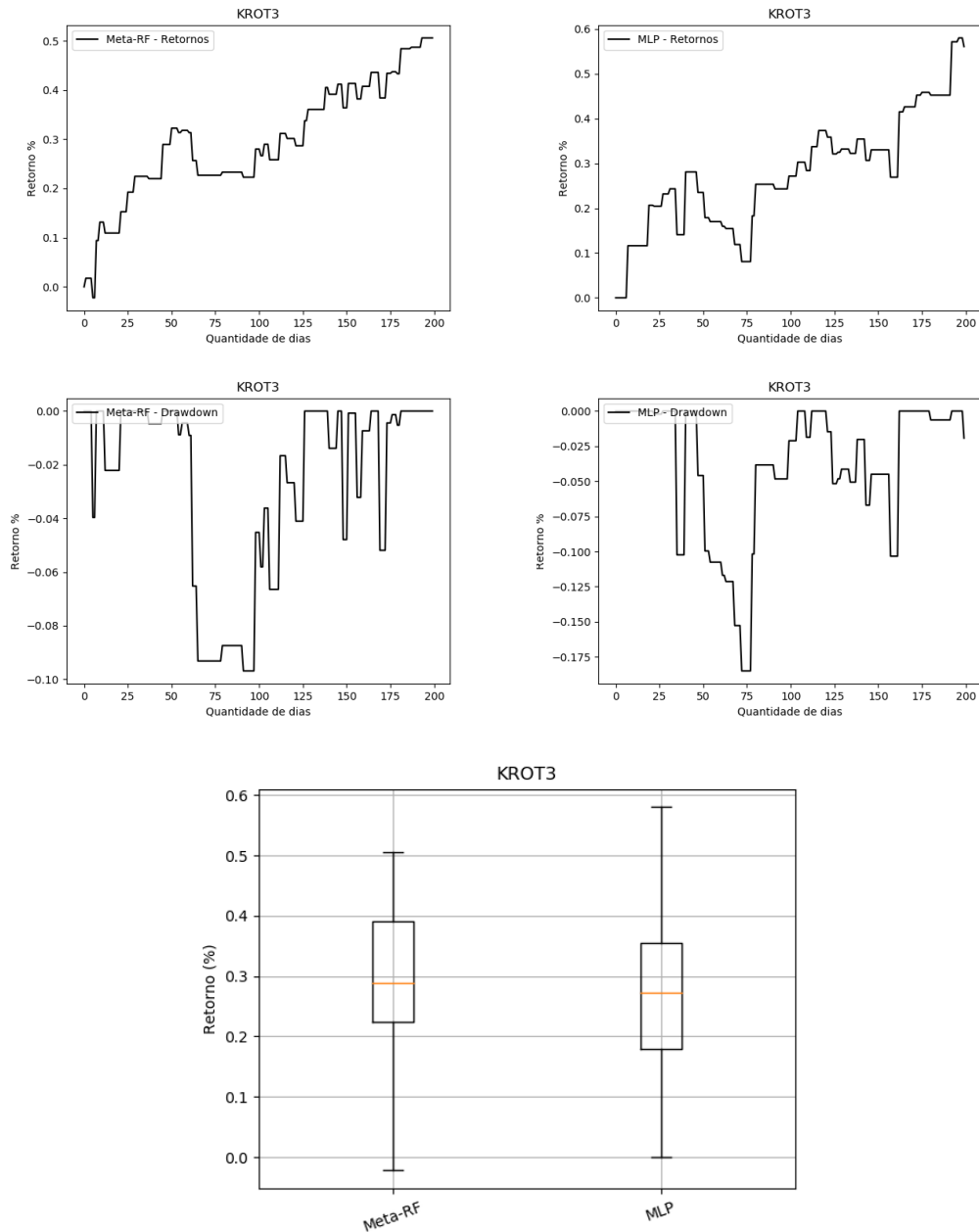
APÊNDICE H – GGBR4

Figura 28 – Comparação Drawdown e % Retornos Financeiros - GGBR4.



APÊNDICE I – KROT3

Figura 29 – Comparação Drawdown e % Retornos Financeiros - KROT3.



APÊNDICE J – GOLL4

Figura 30 – Comparação Drawdown e % Retornos Financeiros - GOLL4.

