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

UM ARCABOUÇO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA AVALIAÇÃO DE ESTRATÉGIAS DE INVESTIMENTO PARA O MERCADO DE AÇÕES

EDUARDO JABBUR MACHADO

Orientador: Adriano César Machado Pereira
CEFET-MG - PPGMMC

BELO HORIZONTE
ABRIL DE 2020

EDUARDO JABBUR MACHADO

UM ARCABOUÇO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA AVALIAÇÃO DE ESTRATÉGIAS DE INVESTIMENTO PARA O MERCADO DE AÇÕES

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: Sistemas Inteligentes

Orientador: Adriano César Machado Pereira
CEFET-MG - PPGMMC

Machado, Eduardo Jabbur
M149a Um arcabouço de algoritmos de aprendizado de máquina para avaliação de estratégias de investimento para o mercado de ações / Eduardo Jabbur Machado. – 2020.
84 f.

Tese de doutorado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.
Orientador: Adriano César Machado Pereira.
Tese (doutorado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Bolsa de valores – Modelagem matemática – Teses. 2. Negociação (Administração de empresas) – Teses. 4. Algoritmos – Teses. 5. Aprendizado de máquina – Teses. 6. Administração financeira – Teses. 7. Administração de risco financeiro – Teses. I. Pereira, Adriano César Machado. II. Centro Federal de Educação Tecnológica de Minas Gerais. III. Título.

CDD 519.6

Elaboração da ficha catalográfica pela bibliotecária Jane Marangon Duarte, CRB 6ª 1592 / Cefet/MG

Dedico este trabalho à minha esposa Ana Paula e à minha filha Maria Eduarda, para lembrá-las que em suas liberdades haverá a opção do caminho do conhecimento.

Agradecimentos

Agradeço primeiramente à Deus pela vida, pelas oportunidades, por me guiar, iluminar e me dar força para alcançar meus objetivos.

À Ana Paula, esposa e mãe dedicada, pela compreensão nas muitas vezes em que precisei dedicar o tempo aos estudos, quando poderia estar com ela e nossa filha Maria Eduarda.

Aos meus pais Antônio e Mércia pelos exemplos e por me proporcionarem o privilégio e a experiência única de crescer em uma família com 5 filhos cheios de alterações.

Aos meus 6 irmãos Adriana, Fernanda, Alexandre, Gustavo, Katarina e Vitor pela nossa história, pelo incentivo e apoio. Tenho muito orgulho de cada um de vocês.

Ao meu orientador, Professor Dr. Adriano César Machado Pereira, pela sua constante ajuda e encorajamento, por compartilhar sua sabedoria, pela confiança depositada, pela paciência, pelos conselhos e por me ajudar a caminhar pela estrada do conhecimento.

Agradeço aos professores participantes da banca examinadora neste momento tão importante e esperado: Dr. Anísio Mendes Lacerda, Dr. Felipe Dias Paiva, Dr. Daniel Hasan Dalip, Dr. Juliano Lima Pinheiro e Dr. Arthur Rodrigo Bosco de Magalhães.

Ao grupo de Finanças Computacionais (FICO) CEFET-MG / UFMG pela oportunidade de trocar experiências, pois praticamente todas as etapas desta pesquisa foram influenciadas pelas ricas contribuições e discussões durante os nossos encontros semanais.

Aos amigos Luiz Ângelo, Reginaldo, Rodrigo, Emerson, Fábio Elias, Carlos Assis e Renato pelos ensinamentos, desabafos e conselhos. Muitas das minhas dificuldades e frustrações foram superadas conversando com vocês.

À CAPES pelo apoio financeiro parcial e ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, corpo docente e discente, pelos 10 anos de muito aprendizado.

Aos familiares, amigos e colegas de trabalho gostaria de desculpar-me pelas ausências e abdicação nos eventos, intercalando o tempo entre estudos, missões e lazer, pois como todos sabem, continuo "...lavando muita caixa d'água...". Forte abraço a todos.

"Ninguém baterá tão forte quanto a vida. Porém, não se trata de quão forte pode bater, se trata de quão forte pode ser atingido e continuar seguindo em frente. É assim que a vitória é conquistada."(Rocky Balboa, 1976).

Resumo

O mercado de ações tem crescido a cada dia e desempenhado um importante papel entre as pessoas físicas que querem investir e as empresas que precisam capitalizar-se para expandir o seu negócio. Porém, devido a grandes variações e oscilações neste tipo de mercado, os investidores estão sujeitos a riscos que podem acarretar também grandes prejuízos. Com o objetivo de estudar os principais trabalhos que envolveram aplicações no mercado de ações relativos aos temas tratamento de dados, modelos de previsão de tendências e modelos de operação, foi realizada uma Revisão Sistemática da Literatura (RSL). Verificou-se que o algoritmo Máquina Restrita de *Boltzmann (RBM)* de categoria não supervisionada de *deep learning* está sendo muito empregado para realizar redução de dimensionalidade através da seleção e extração de características latentes dos dados de séries temporais. Com o objteivo otimizar o tempo gasto em pesquisas acadêmicas na implementação de algoritmos e modelos por estudantes e interessados da área de mercado financeiro e, para que possam dedicar esforços em criar, validar e aprimorar suas estratégias de investimeno, propõe-se a implementação de um arcabouço automatizado constituído de 5 etapas: Extração de Dados, Transformação de Dados, Classificação de Modelos de Previsão de Tendências, Estratégia de Operação e Análise de Resultados. Durante as simulações, avaliaram-se dados de cotações históricas de 50 ativos negociados na Bolsa Brasil Balcão (B3), acrescidos de mais 10 índices globais, para um período de 1.000 dias úteis (2010 a 2014) na etapa de Treinamento e outros 1.000 dias úteis (2015 a 2019) na etapa de Validação, para os 8 modelos de previsão de tendências propostos. Como resultado foi selecionado o modelo combinado *RBM + NB* que gerou para dos 50 ativos da B3 juntamente com os 10 índices globais uma Precisão média de 52,85%, uma Acurácia média de 53,11% e um Retorno Financeiro Bruto médio de 50,93%, sendo 28,02% superior ao *baseline* de *Buy and Hold* médio de 30,60%.

Palavras-chave: Caracterização e transformação de dados. Mercado de ações. Estratégias de negociação. Revisão Sistemática da Literatura. Algoritmos de aprendizado de máquina. Aprendizado profundo. Medidas de desempenho. Medidas de risco.

Abstract

The capital market is growing every day and playing an important role among individuals who want to invest their money and companies that need capitalize to expand their business. However, due to large variations and fluctuations in this type of market, investors are subject to risks which can result in many losses. A Systematic Literature Review (SLR) was carried out in order to collect the main works involving applications in the stock market related to the following data processing, trend forecasting models and operation models. It was found that the Restricted Boltzmann Machine (RBM) algorithm of deep learning is being widely used to perform dimensionality reduction through the selection and extraction of latent characteristics from time series data. In order to avoid that other students and interested in the financial market area have to spend a lot of time on their research in the implementation of algorithms and can dedicate efforts in creating, validating and improving their trading strategies, the implementation of an automated framework consisting of 5 steps (Data Extraction, Data Transformation, Classification of Trend Forecasting Models, Operation Strategy and Analysis of Results). During the simulations, the historical quotation data of 50 assets traded on the Brazilian Stock Exchange increases by more than 10 global indices, for a period of 1,000 working days (2010 to 2014) in the training stage and another 1,000 working days (2015 to 2019) in the validation stage, for the 8 proposed trend forecasting models. As a result, the combined model RBM+NB that generated for Brazilian Stock Exchange 50 assets was selected together with the 10 global indexes an average Accuracy of 52.85%, an average Accuracy of 53.11% and an average Gross Financial Return of 50.93%, 28.02% being higher than the Buy and Hold average of 30.60% of baseline.

***Keywords**: Data Characterization. Data transformation. Stock Market. Negotiation strategies. Systematic Literature Review. Machine Learning Algorithms. Deep Learning. Performance Measures. Risk Measures.*

Lista de Figuras

Figura 1 – Etapas da Revisão Sistemática da Literatura 7
Figura 2 – Resumo da Fundamentação Teórica utilizada: Mercado de Ações, Aprendizado de Máquina e Métricas de Avaliação. 29
Figura 3 – Arcabouço proposto. 30
Figura 4 – Treinamento: Avaliação dos 8 modelos propostos para os 50 ativos, período de 1.000 dias úteis de 04/11/2010 a 13/11/2014. 45
Figura 5 – Comparação entre o custo operacional, o retorno bruto e o retorno líquido acumulado de 132 dias úteis de 01/12/2014 a 16/07/2015. 46
Figura 6 – Validação: Avaliação dos 8 modelos propostos para os 50 ativos, período de 1.000 dias úteis de 05/01/2015 18/11/2018. 47
Figura 7 – Validação dos 10 índices globais: comparação entre as medidas de desempenho e o retorno financeiro de 1.000 dias úteis de 05/01/2015 18/11/2018. 52
Figura 8 – Séries de Preço de Fehamento: 10 índices globais, 1.000 dias úteis (07/01/2015 a 24/01/2019). 55
Figura 9 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 10 índices globais, 1.000 dias úteis (07/01/2015 a 24/01/2019). 56
Figura 10 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 1 de 5 68
Figura 11 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 2 de 5 69
Figura 12 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 3 de 5. 70
Figura 13 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 4 de 5. 71
Figura 14 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 5 de 5. 72
Figura 15 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 1 de 5. 74
Figura 16 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 2 de 5. 75

Figura 17 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 3 de 5. 76
Figura 18 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 4 de 5. 77
Figura 19 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 5 de 5. 78

Lista de Tabelas

Tabela 1 – Variação do CDI e do Ibovespa nos últimos 9 anos. 1
Tabela 2 – Quantitativo de estudos coletados 11
Tabela 3 – Comparação entre os trabalhos relacionados selecionados pela RSL. . 17
Tabela 4 – Exemplo de modelo original da série histórica de preços. 31
Tabela 5 – Exemplo de modelo normalizado da série histórica de preços. 32
Tabela 6 – Variação acumulada de indicadores de *Baselines* do mercado brasileiro. 39
Tabela 7 – Descrição dos 50 ativos da B3 que serão utilizados para as fases de Treinamento (2013 a 2014) e Validação (2015 a 2018) dos modelos de previsão de tendências. 40
Tabela 8 – Intervalo de parâmetros de calibração dos algoritmos. 42
Tabela 9 – Fase de Treinamento: 50 ativos - Retorno Financeiro Bruto (%) - Critério de seleção entre os 8 modelos de previsão de tendências propostos. . . 44
Tabela 10 – Valores utilizados e exemplos dos custos operacionais 46
Tabela 11 – Fase de Validação: 50 ativos - Retorno Financeiro Bruto (%) - Critério de seleção entre os 8 modelos de previsão de tendências propostos. . . . 48
Tabela 12 – Modelo RBM + NB: Treinamento (1.000 dias de 04/11/2010 a 13/11/2014) e Validação (1.000 dias de 07/01/2015 a 24/01/2019). 49
Tabela 13 – Validação: Descrição dos 10 índices globais por setor, *buy and hold*, período de Validação (2015 a 2018) e a distribuição de classes de (Baixa e Alta). 51
Tabela 14 – Validação: 10 índices globais - Retorno Financeiro Bruto (%) - Critério de seleção entre os 8 modelos de previsão de tendências propostos. . . . 51
Tabela 15 – Extrato consolidado da Validação do modelo *RBM + KNN* para os 10 índices globais no período de 1.000 dias (07/01/2015 a 24/01/2019). . . 53
Tabela 16 – validação: modelo *RBM + NB* - Resultado da média das simulações. . . 57
Tabela 17 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 1 de 5. 80
Tabela 18 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 2 de 5. 81
Tabela 19 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 3 de 5. 82
Tabela 20 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 4 de 5. 83
Tabela 21 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 5 de 5. 84

Lista de Algoritmos

Algoritmo 1 – Pseudocódigo da Estratégia principal 34

Algoritmo 2 – Pseudocódigo da estratégia Comprar Sempre 34

Algoritmo 3 – Pseudocódigo da estratégia *Buy and Hold* 35

Algoritmo 4 – Pseudocódigo da estratégia aleatória 35

Algoritmo 5 – Pseudocódigo das estratégias Alvos 0,50%, 1,00% e 2,00% 36

Algoritmo 6 – Pseudocódigo da estratégia inversa 36

Lista de Abreviaturas e Siglas

AD 48	Árvore de Decisão
AR	Modelos Autorregressivos
ARIMA	Autoregressivo de Média Móvel
BD	Base de Dados
B3	Brasil Bolsa Balcão
BP	*Backpropagation*
CB	Critérios de Busca
CDI	Certificados de Depósitos Interbancários
CE	Critérios de Exclusão
CI	Critérios de Inclusão
CNN	*Convolutional Neural Network*
CQ	Critérios de Qualidade
CS	Critérios de Seleção
DARF	Documento de Arrecadação de Receitas Federais
DBN	*Deep Belief Network*
DC	Dados Coletados
DNN	*Deep Neural Networks*
EQM	Erro Quadrático Médio
FNN	Feedforward Neural Network
GARCH	Autoregressivo não estacionário
Ibovespa	Índice Bovespa
IIM3	Instituto Internacional de Meteorologia
IR	Imposto de Renda

ISS	Imposto Sobre Serviços
KNN	*K-Nearest Neighbor*
LASSO	Modelo de Regressão Linear
LSTM	*Long Short-Term Memory*
MLP	Multilayer Perceptrom
NB	*Naive Bayes*
PCA	Análise dos Componentes Principais
QP	Questão de pesquisa
RBM	*Restricted Boltzmann Machine*
RF	*Random Forest*
RIDGE	Modelo de Regressão Linear
RL	Regressão Logística
RMSE	Raiz do Erro Quadrado Médio
RNA	Rede Neural Artificial
RNN	*Recurrent Neural Network*
RSL	Revisão Sistemática da Literatura
Selic	Sistema Especial de Liquidação e Custódia
SVM	Máquina de Suporte de Vetores
SVR	*Support Vector Regression*

Sumário

1 – Introdução 1
1.1 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 Revisão Sistemática da Literatura 6
2.1.1 Planejamento 6
2.1.2 Realização 8
2.1.3 Apresentação de Resultados 11
2.2 Trabalhos Relacionados 11

3 – Fundamentação Teórica 19
3.1 Introdução ao Mercado de Ações 19
3.1.1 Análise Técnica 20
3.1.2 Estratégias de negociação 22
3.1.3 Custos Operacionais 22
3.2 Algoritmos de Aprendizado de Máquina 23
3.2.1 Árvore de Decisão (AD J48) 24
3.2.2 *Naive Bayes* (NB) 25
3.2.3 Máquinas de Suporte de Vetores (SVM) 25
3.2.4 *K-Nearest Neighbor (KNN)* 26
3.2.5 Máquina Restrita de *Boltzmann (RBM)* 26
3.3 Métricas de Avaliação 27
3.4 Considerações Finais 29

4 – Metodologia 30
4.1 Extração de Dados 31
4.2 Transformação de Dados 31
4.3 Modelos de Previsão de Tendência 33
4.4 Estratégias de Operação 33
4.5 Análise de Resultados 37
4.6 Considerações finais 38

5 – Experimentos: Resultados e Análises 39

5.1 Simulação do Treinamento dos Modelos 41
5.1.1 Avaliação da calibração dos algoritmos 41
5.1.2 Avaliação da janela de classificação 42
5.1.3 Avaliação dos modelos de previsão de tendências 43
5.1.4 Avaliação da viabilidade dos custos operacionais 45
5.2 Simulação da Validação dos Modelos 47
5.2.1 Avaliação de 10 índices globais 51
5.3 Considerações Finais 57

6 – Conclusão 58
6.1 Limitações e escopo 60
6.2 Publicações realizadas 61
6.3 Trabalhos futuros 61

Referências 62

Apêndices 66

APÊNDICE A – 50 ativos da B3: Gráficos de séries de Preço de Fechamento . . 67

APÊNDICE B – 50 ativos da B3: Gráficos de Retornos Acumulados 73

APÊNDICE C – 50 ativos da B3: Resultados detalhados 79

Capítulo 1

Introdução

O Brasil sempre foi um país de juros elevados e a maioria de seus investidores, por questões culturais, alocavam grande parte de seus recursos em aplicações de renda fixa como a poupança, certificados de depósito bancário (CDB), tesouro direto, dentre outros (EXAME, 2019b).

O Certificado de Depósito Interbancário (CDI) e o Sistema Especial de Liquidação e Custódia (SELIC) são indexadores de investimentos baseados em taxas de juros. Ambas são muito importantes na economia do país e seus valores andam lado a lado. A SELIC é a principal taxa de juros da economia brasileira e é conhecida como uma taxa de juros pública utilizada pelo governo para remunerar aqueles bancos que emprestam dinheiro entre si utilizando-se de títulos do Tesouro Nacional (LANGONI, 2019).

O CDI tem a mesma função, mas se refere aos empréstimos de curto prazo realizados entre os próprios bancos que se utilizam dos próprios recursos dos bancos ao invés de um título do Tesouro Nacional. Com o passar do tempo, as taxas praticadas no CDI passaram a servir de referência para o mercado de crédito e, consequentemente, para o próprio mercado financeiro. Como os bancos conseguem emprestar dinheiro praticamente sem risco no mercado interbancário, o CDI acabou virando um padrão para analisar qualquer operação de investimento no mercado de renda fixa.

Tabela 1 – Variação do CDI e do Ibovespa nos últimos 9 anos.

Índices / Ano	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
CDI (%)	11,59	8,40	8,06	10,81	13,24	14,00	9,93	6,42	4,66
Ibovespa (%)	-18,11	7,40	-15,50	-2,91	-13,31	38,93	26,86	15,03	31,58

Fonte dados coletados no site da Cetip (CETIP, 2020).

No entanto, com as sucessivas reduções da taxa SELIC nos últimos 3 anos, impactando diretamente no valor do CDI, conforme a Tabela tab:cdi, muitos investidores do mercado de

renda fixa passaram a procurar outras opções no mercado de renda variável, como uma alternativa de investimento, visando garantir uma maior rentabilidade para seu capital.

De acordo com a Brasil Bolsa Balcão (B3), o número de investidores pessoa física, saltou de 619.625 em 2017 para 1.830.745 em 2019, gerando um aumento de 295,46% (B3, 2020). Junto com esta demanda, aumentou também na internet a quantidade de serviços (análises, capacitação, consultoria, dicas), por pessoas que não são certificadas e habilitadas para atuar no mercado de ações, com a promessa de obter ganhos significativos em curtíssimo prazo (EXAME, 2019a).

O mercado de ações tem a característica de ser sensível a influências externas, que podem provocar flutuações no valor atual dos ativos. Entre os principais acontecimentos políticos, econômicos e sociais, que ocorreram no Brasil nos últimos 4 anos (2015 a 2019), destacam-se, de forma pontual, os seguintes eventos que impactaram de forma direta ou indireta:

- Política: investigação, delações, prisões, impeachment, reformas e eleições;
- Epidemias: casos de Dengue, Chikungunya e Zika vírus;
- Crise econômica: medidas de austeridade e limitações de gastos;
- Crise no Sistema Prisional: rebeliões em quase todos os Estados;
- Greve de caminhoneiros: aumento abusivo, obstrução, falta e perda de mercadorias;
- Minas Gerais: desastres, impacto ambiental e social nas regiões devastadas;
- Rio de Janeiro: Intervenção militar e Jogos olímpicos;
- Crise humanitária na Venezuela: aumento da imigração, falta de serviços básicos; e
- Desmatamento: aumento na região amazônica, dentre outros.

Apesar de vários indicativos técnicos para uma determinada tomada de decisão, o investidor ainda sim age na direção contrária, devido questões emocionais como impulso, razões, medos e vontade de ganhar que podem interferir em suas decisões, valorizando demais alguns ganhos e menosprezando perdas.

Diante do aumento do número de investidores pessoas físicas na B3, da dificuldade para pesquisar, filtrar e analisar o grande volume informações disponíveis e para eliminar fatores emocionais em decisões de investimento, surge a necessidade de desenvolver uma forma automatizada para a avaliação de modelos de operação de estratégias para o mercado de ações.

1.1 Problema de Pesquisa

O problema de pesquisa desta tese é a falta de um modelo de operação automatizado para investir no mercado de ações. Diante deste cenário, o problema de pesquisa foi decomposto

em três questões principais a serem investigadas:

- Tratamento dos Dados: devido à grande quantidade de fontes de dados distintas disponíveis na Web, 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: Quais técnicas de ciências dos dados são necessárias e adequadas para selecionar, preparar e analisar dados de séries temporais financeiras?
- Modelo de Previsão: a literatura possui muitos algoritmos de classificação que podem ser parametrizados com várias configurações, dificultando diretamente na qualidade do aprendizado do modelo de previsão. Com base neste problema, surge a seguinte questão: Como configurar, treinar, medir, combinar e selecionar um ou vários algoritmos de classificação para a geração de previsão de tendências que garanta uma taxa mínima de aprendizado em relação à sua precisão como medida de desempenho?
- Estratégias de Operação: é um grande desafio modelar estratégias de operação 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 no mercado de ações?

Esta pesquisa é composta de um tema interdisciplinar que envolve algumas áreas do conhecimento, como por exemplo, matemática, estatística, economia, finanças, ciência de dados e inteligência computacional de forma integrada e aplicada.

1.2 Objetivos: Geral e Específicos

O objetivo geral desta pesquisa é realizar a modelagem e implementação de um arcabouço automatizado, que utiliza séries temporais financeiras de cotações históricas de ativos da B3 e índices globais, como entrada de algoritmos de aprendizado de máquina, que são utilizados como classificadores, para realizar de previsão de tendências, com a finalidade de avaliar estratégias de investimentos para o mercado de ações.

Os objetivos específicos norteiam o entendimento dos cenários de estratégias de negociação, simulação e análise de resultados para:

- Modelar e implementar um arcabouço automatizado composto por 5 etapas (Extração de Dados, Caracterização e Transformação de Dados, Avaliação de Modelos de Previsão de Tendências, Avaliação de Estratégias de Operação e Análises de Resultados);

- Analisar as fontes de dados disponíveis para caracterizar, transformar, normalizar e criar novos indicadores, para a integração dos atributos que serão utilizados nas simulações;
- Avaliar os algoritmos de classificação tradicionais (Árvore de Decisão (AD J48), *Naive Bayes* (NB), Máquinas de Suporte de Vetores (SVM) e *K-Nearest Neighbor* (KNN)) de forma isolados e combinados com o algoritmo de aprendizado profundo Máquina Restrita de Boltzmann (RBM), em relação às medidas de desempenho Precisão (%) e ao retorno financeiro (%);
- Modelar 8 estratégias de investimento como *baselines* de comparação em relação ao percentual de retorno financeiro acumulado, considerando incluir e ou retirar atributos tanto no modelo de previsão, quanto na estratégia de investimento; e
- Realizar simulações e *backtesting* dos modelos de previsão propostos validando-os através de um simulador realístico do mercado financeiro contemplando as variáveis de custos operacionais.

1.3 Contribuições

A principal contribuição deste trabalho é a implementação de um arcabouço automatizado composto por 5 etapas (coleta, transformação, classificação, operação e análise de resultados), que utiliza dados integrados de fontes distintas de cotações históricas como entrada de modelos de classificação para previsão de tendências, combinando algoritmos de aprendizado de máquina, através de estratégias de negociação avaliadas em relação à análise de indicadores de dados estatísticos, medidas de desempenho, operação e risco, para a sociedade acadêmica.

A disponibilização deste arcabouço contribui para que outros estudantes e interessados no mercado de ações minimizem o tempo gasto em suas pesquisas na implementação de algoritmos e, possam dedicar esforços em criar, validar e aprimorar suas estratégias de investimento.

Além do contexto de mercado de ações, o arcabouço automatizado implementado como solução proposta, desde que adaptado em relação aos dados e objetivos, pode ser utilizado em outras áreas, como por exemplo: os mercados (Dólar, Índices e Criptomoedas), análise de crédito para caracterização de clientes, detecção de fraude em transações de cartões de crédito, análise de diagnósticos médicos, análise de propagação de epidemias, análise de reconhecimento de imagens, dentre outras.

1.4 Organização do Trabalho

O restante do trabalho está organizado da seguinte maneira. O Capítulo 2 apresenta uma Revisão Sistemática da Literatura para selecionar e avaliar os trabalhos relacionados. O Capítulo 3 descreve a fundamentação teórica, que é baseada em análise técnica com algumas teorias econômicas, indicadores financeiros e conceitos de aprendizado de máquina. O Capítulo 4 descreve a metologia aplicada neste trabalho, detalhando cada etapa a ser realizada. O Capítulo 5 descreve as configurações e os experimentos realizados a fim de calibrar e selecionar o tipo de dado e o modelo de previsão de tendências que melhor se adapta ao cenário avaliado. Serão apresentados também os resultados e análises da simulação realizada contemplando 50 ativos da B3 e 10 principais índices globais. Finalmente, o Capítulo 6 apresenta a conclusão, as limitações e dificuldades encontradas, questões de estudo ainda em aberto, a listagem das publicações realizadas e, por último, os trabalhos futuros.

Capítulo 2

Revisão da Literatura

O Capítulo 2 trata do levantamento das iniciativas já existentes sobre desenvolvimento de estratégias de negociação baseadas em algoritmos de aprendizado de máquina para a aplicação no mercado de ações. Com o propósito de identificar e extrair conceitos, fundamentos e trabalhos relacionados de uma forma metodológica, realizou-se uma revisão sistemática da literatura, adotando o conjunto de diretrizes propostas por (KITCHENHAM, 2007).

2.1 Revisão Sistemática da Literatura

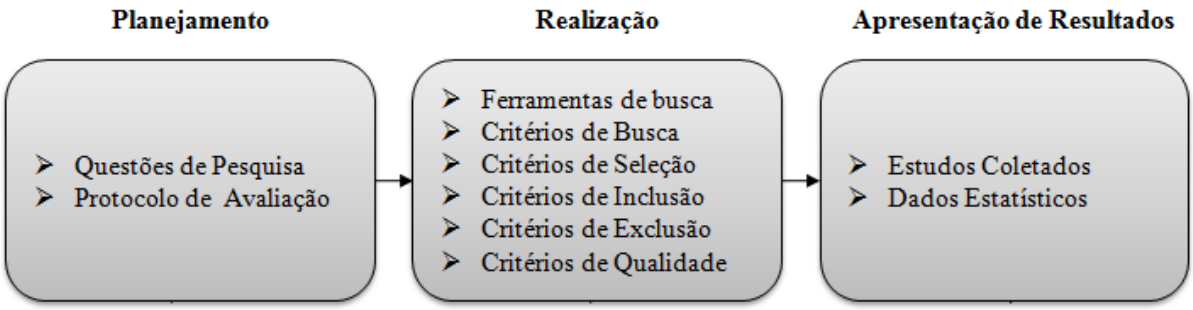
(KITCHENHAM, 2007) define a revisão sistemática da literatura (RSL) como um meio de identificar, avaliar e interpretar uma grande parte das pesquisas disponíveis relevantes para a questão de pesquisa específica, área temática ou fenômeno de interesse. Uma RSL possui três etapas principais: planejamento, realização e apresentação dos resultados, conforme Figura 1.

A principal característica que diferencia uma revisão da literatura de uma RSL, é que esta realiza a definição do protocolo de avaliação na etapa de planejamento, que trata os dados coletados como pré-requisito para uma avaliação posterior. A seguir serão detalhadas cada etapa juntamente com seus procedimentos realizados durante o processo RSL aplicado ao trabalho em questão.

2.1.1 Planejamento

A etapa de planejamento identifica a necessidade da revisão através do levantamento das questões de pesquisa para o desenvolvimento do protocolo de avaliação (KITCHENHAM, 2007), sendo necessário ter a definição do tema principal do trabalho que será desenvolvido, juntamente com a hipótese que o pesquisador objetiva corroborar.

Figura 1 – Etapas da Revisão Sistemática da Literatura



Esta proposta de pesquisa tem como tema principal o desenvolvimento de estratégias de negociação baseadas em técnicas de aprendizado de máquina com o objetivo de realizar caracterização, classificação e previsão de dados para futuras aplicações no mercado de financeiro.

Um trabalho inicial de pesquisa das possíveis palavras chaves que fazem parte do contexto do tema escolhido, como critério de busca, se faz necessário para apoiar a construção e direcionamento das questões de pesquisas específicas. Em seguida, definem-se as questões de pesquisas (QP) decompostas do tema principal:

- **QP1:** Quais as estratégias / algoritmos de negociação existentes para o mercado de ações?
- **QP2:** Quais são as teorias econômicas existentes na literatura para o mercado de ações?
- **QP3:** Como aplicar teorias econômicas existentes na literatura utilizando técnicas de *Machine Learning* e *Deep Learning* para propor estratégias de negociação para o mercado de ações?
- **QP4**: Quais as bases e ou fontes de dados disponíveis e como realizar a extração / coleta destes dados?
- **QP5**: Como tratar o dado bruto, aplicação de normalização, técnicas de agrupamento e redução de dimensionalidade?
- **QP6**: Quais técnicas e ou modelos mais adequados, suas configurações e parâmetros?
- **QP7**: O que já foi realizado, quais as técnicas e ou modelos mais adequados e os resultados iniciais / parciais?
- **QP8**: Quais as medidas de desempenho e de avaliação de modelos de predição?

As atividades são iniciadas durante a fase de desenvolvimento do protocolo, que é subdividida em 6 (seis) partes distintas propostas por (KITCHENHAM, 2007), que serão detalhadas a seguir:

- **Base de dados e ferramentas de busca:** é necessário um estudo prévio sobre o tema do trabalho que será abordado, com o objetivo de catalogar as possíveis bases de dados candidatas, que possam fornecer trabalhos relacionados em relação ao projeto proposto;
- **Critérios de busca:** após a definição do tema principal e das questões de pesquisas, são relacionadas as possíveis palavras chaves que fazem parte do contexto do trabalho;
- **Critérios de inclusão:** é a definição de quais tipos de trabalhos (artigos, dissertações, dentre outros), que serão coletados e durante o critério de busca da revisão sistemática da literatura;
- **Critérios de exclusão:** é constituído de uma análise qualitativa dos trabalhos coletados, em que 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:** é o refinamento de todos os trabalhos coletados, com o objetivo de selecionar por categoria os trabalhos mais relevantes que identificam-se com o propósito do tema da pesquisa; e
- **Dados coletados:** após a triagem da etapa anterior, as informações dos trabalhos coletados são catalogadas para posterior consulta em relação aos parâmetros de publicação, como por exemplos, ano, tipo, título, autores e etc, com o foco de tratar categorizar os trabalhos selecionados.

A seleção dos estudos primários se rege por critérios de inclusão e exclusão, sendo inicialmente especificados na definição do protocolo, e podem ser refinados a medida em que os critérios de qualidade são definidos. A forma de extração dos dados ou critérios de busca são preparados durante a construção do protocolo e alteradas quando são definidos os critérios de qualidade. Os parâmetros para síntese de dados definidos inicialmente no protocolo podem ser alterados na medida em que os dados são coletados.

2.1.2 Realização

O processo de seleção e filtragem dos estudos primários é iniciado conforme os critérios estabelecidos no protocolo de avaliação. Nesta fase, o processo seleciona os estudos primários e verifica a qualidade dos trabalhos, coleta os dados e realiza a síntese dos resultados. Realizou-se o processo de busca das pesquisas durante o período de 2016 a 2018, onde vários trabalhos foram coletados, avaliados, selecionados e eventualmente descartados, devidos não aderirem mais ao tema proposto e ou ao surgimento de novos estudos a todo momento durante o desenvolvimento desta pesquisa.

Base de dados e ferramentas de busca (BD):

- BD1: www.files.scire.net.br;
- BD2: www.mmc.cefetmg.br;
- BD3: www.dl.acm.org;
- BD4: www.hcibib.org;
- BD5: www.sciencedirect.com;
- BD6: www.ieeexplore.ieee.org;
- BD7: www.scholar.google.com.br; e
- BD8: www.google.com.br

Critérios de busca (CB):

- CB1: aprendizado de máquina **AND** (algoritmos de negociação * **OR** mercado de ações * **OR** teorias econômicas * **OR** estratégias de negociação * **OR** agentes inteligentes * **OR** negociação em alta frequência * **OR** Princípio das Ondas de *Elliott* * **OR** caracterização de dados * **OR** indicadores financeiros * **OR** *candlestick* * **OR** *day trade*); e
- CB2: *deep learning* **AND** (*algorithmic trading* * **OR** *stock market* * **OR** *economic theory* * **OR** *strategies trading* * **OR** *intelligent agents* * **OR** *high frequency trading* * **OR** *Elliott Wave Principle* * **OR** *caracterization data* * **OR** *financial indicators* * **OR** *candlestick* * **OR** *day trade*).

Critérios de Seleção (CS):

- CS1: Leitura do título dos trabalhos recuperados a fim de se obter os artigos relevantes e excluir estudos que não sejam relevantes;
- CS2: Leitura dos resumos e palavras-chave dos artigos selecionados com o objetivo de verificar se o trabalho está relacionado com as questões de pesquisa, caso contrário ele será excluído dos conjunto de trabalhos relevantes;
- CS3: Leitura superficial, iniciando pela introdução, tópicos e conclusão dos artigos para verificar se o referido trabalho, realmente está relacionado com as questões de pesquisa; e
- CS4: Leitura completa dos artigos selecionados para avaliação da qualidade e extração dos dados.

Critérios de Inclusão (CI):

- 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; e

- CI6: capítulos de livros.

Critérios de Exclusão (CE):

- CE1: Desprezar trabalhos que não disponibilizam o texto completo para leitura;
- CE2: Trabalhos repetidos e semelhantes; e
- CE3: Trabalhos que não abordam o tema proposto.

Dados Coletados **Critérios de Qualidade (CQ):**

- CQ1: Foi realizada a caracterização do problema?
- CQ2: Os objetivos geral e específicos estão explícitos?
- CQ3: Está descrito como os dados foram coletados?
- CQ4: Está descrito como os dados foram processados?
- CQ5: Está descrito como os dados foram analisados?
- CQ6: É possível reproduzir o estudo em questão?

Dados Coletados (DC):

- DC1: Título do Trabalho;
- DC2: Autoria;
- DC3: Ano;
- DC4: Local Publicação;
- DC5: País da Instituição dos Autores;
- DC6: Área do Conhecimento do veículo de publicação;
- DC7: Palavras-chave; e
- DC8: Resumo do trabalho.

Em seguida é realizada a identificação das possíveis palavras chaves candidatas para acesso em cada uma das bases de dados e ferramentas de busca citadas anteriormente, com base em conhecimento prévio de trabalhos e termos que englobam o cenário do mercado de ações.

Durante a execução do protocolo de pesquisa é necessária a leitura de parte do trabalho analisado em questão, no entanto, tem-se a opção de descartar um trabalho não relevante ao tema sem a seja que realizada a leitura completa deste material. Caso ocorra alguma dúvida em relação à relevância do trabalho analisado, a sequência proposta nos critérios de seleção (CS) será realizada.

Após a execução do protocolo de pesquisa proposto, as publicações selecionadas foram catalogadas usando a ferramenta *JabRef*[1] responsável por organizar e facilitar a catalogação

[1] O JabRef é um gerenciador de referências, baseado em bases de dados *BibTex* e programado em *JAVA*, que realiza a gestão de vários tipos de referências (livros, artigos, anais, etc), e armazena metadados das

e posterior consulta aos estudos selecionados.

2.1.3 Apresentação de Resultados

A Tabela 2 apresenta o quantitativo dos trabalhos coletados em cada uma das etapas da fase de seleção de estudos. A coluna denominada Inicial corresponde à quantidade de trabalhos selecionados que foram obtidos através da *string* de busca em cada uma das bases de dados. As próximas colunas exibem a quantidade de trabalhos selecionados após a aplicação dos critérios de inclusão e exclusão dos trabalhos, conforme os critérios de seleção (CS) já descritos na etapa de realização da RSL. Finalmente, a última coluna corresponde ao número total final de estudos efetivamente selecionados, que serão descritos na Seção 2.2 como trabalhos relacionados.

Tabela 2 – Quantitativo de estudos coletados

Base de dados	**Inicial**	**Etapa 1**	**Etapa 2**	**Etapa 3**	**Etapa 4**
files.scire /mmc.cefetmg	32	20	7	3	2
dl.acm.org	29	22	19	13	6
hcibib.org	19	12	7	4	3
sciencedirect.com	61	58	30	25	5
ieeexplore.ieee.org	31	28	23	16	4
google scholar	38	36	14	12	6
Total	268	230	152	114	26

Após o mapeamento apresentado do cenário atual através da aplicação da revisão sistemática da literatura, visualiza-se como uma oportunidade real de aplicação deste tema em um projeto de doutorado, com o objetivo de desenvolver novos caminhos, estudos e pesquisas com várias possibilidades de aplicações para propor estratégias de negociação para o mercado de ações.

2.2 Trabalhos Relacionados

Nesta seção, serão apresentadas as principais pesquisas selecionadas pela revisão sistemática da literatura com o detalhamento de alguns estudos que abordam os temas de análise de informações da web e aplicações de aprendizado de máquina.

A popularização da internet com a utilização de dispositivos móveis, como *smartphones*, *tablets* e computador, por meio de sites de notícias, aplicativos e redes sociais são considerados os principais meios de articulação, interação e divulgação de ideias entre a sociedade, organizações e instituições, o que possibilitou um aumento na utilização de novas tecnologias computacionais para comunicação e disseminação da informação, como

publicações podendo inclusive associar arquivos em diversos formatos à referência.

fonte de dados, para a aplicação de estratégias de investimento no contexto de bolsa de valores.

Neste contexto, várias aplicações, ferramentas e plataformas de previsão que contém mensagens de conversas, coletadas em tempo real, sobre as condições do mercado e valores de ativos específicos estão disponíveis atualmente, como por exemplo, o *StockTwits*[2] utilizado em vários trabalhos (GOMIDE, 2012), (FERREIRA, 2016), (SANTOS, 2016). Investidores e outros interessados em ações e mercados, podem seguir facilmente ações individuais, contribuidores específicos, além de visualizar a postagem do ativo em vários sites financeiros que integram as postagens, incluindo *Yahoo!* Finanças, *CNNMoney*, *Reuters*, *TheStreet.com*, *Bing.com* e *The Globe and Mail*. Esta plataforma também é integrada com as principais redes sociais, incluindo *Twitter*, *Facebook* e *LinkedIn*.

A utilização de 3 algoritmos de aprendizado de máquina, Naive Bayes (NB), Regressão Logística (RL) e Support Vector Machines(SVM) aplicados a dados históricos de séries financeiras, foi proposta por (SILVA, 2014) com objetivo de encontrar tendências para os preços de ações do mercado Brasileiro em determinados períodos de tempo. Os dados utilizados foram cotações históricas do Ibovespa com valor diário de fechamento para o período avaliado de 01/01/2006 a 31/12/2012. A fim de obter modelos com maior capacidade de generalização, foi alterado o comportamento do método de validação cruzada padrão para se ajustar às características do problema, tornando-o deslizante em função do tempo. Observou-se que retreinar o modelo em determinado período de tempo melhora positivamente a taxa de acerto quando comparado ao modelo de validação cruzada padrão. Como apresentado em estudos similares da área, o algoritmo SVM apresentou melhores resultados que os algoritmos comparados nesse estudo. Uma análise detalhada mostra que o ajuste de parâmetros traz benefícios para o resultado dos modelos, com ganhos de até 32%.

(VENGERTSEV, 2014) realiza a comparação de técnicas de *machine learnig* e *deep learning* para a previsão de séries temporais, apresentando como solução ótima a técnica de *Restricted Boltzmann Machines (RBM)*. O período avaliado foi de 1428 meses com a coleta de dados demográficos, financeiros, industrial, macro e micro econômicos disponibilizados pelo IIM[3], granularidade diária. Realizou a avaliação de 5 (cinco) modelos Autoregressivo de Média Móvel (ARIMA), *Support Vector Regression* (SVR), *Multilayer Perceptrom* (MLP) e RBM + MLP, onde os dados foram divididos em dois grupos: Treinamento de 80% a 90% e Teste de 20% a 10%. Como resultados o trabalho apresentou a raiz do erro quadrado

[2] *StockTwits*: é uma plataforma de comunicações financeiras para a comunidade financeira e de investimento, consiste em postagens que incluem ideias, *links*, gráficos e outros dados financeiros importantes, resumidos dentro de 140 caracteres. Os usuários são analistas, mídia e investidores de todos os tipos e empresas públicas.

[3] (IIM) Instituto Internacional de Meteorologia, M3 dados de competição disponível em https://forecasters.org/resources/time-series-data/m3-competition

médio (RMSE) de cada um dos 5 modelos nas fases de Treinamento e Teste.

No trabalho de (ESTRADA, 2015) utilizaram-se técnicas de *deep learning* para a previsão de dados financeiros, em especial DBN junto com MLP, para a escolha de ativos em portifólio através do retorno médio do preço dos ativos do índice S&P500. O período avaliado é de 01/01/1985 a 31/12/2006, totalizando 134.350 amostras. Os dados foram separados em 70% para Treinamento, 15% para Validação e 15% para Teste, onde os parâmetros foram escolhidos através de validação cruzada. Os 3 (três) modelos avaliados foram *Gauss-Bernoulli* RBM com a utilização de *Graphics Processing Unit* (GPU) para 100 épocas, DBN-MLP com 3 camadas ocultas em *Deep Belief Network* (DBN) e 1 camada oculta no MLP e Regressão Logística com somente 1 neurônio e MLP contendo 1 camada oculta. Como resultados o trabalho apresentou o erro na validação de 46,52% e no teste de 47,11%.

(ROCA, 2015) coletou dados diários de 19 índices da B3, para o período de 01/01/2006 a 31/01/2011 totalizando 1250 dias, onde foram implementados 2 modelos de RNA contendo dados de entrada diários baseados em preço e outra Rede Neural Artificial (RNA) em retornos diários denotando observações sucessivas de uma variável tomada realizando a transformação da série de valores em uma série de retornos. A fase de treinamento foi de 4 anos e a fase de teste de 1 ano de dados. Utilizou a Análise dos Componentes Principais (PCA) para fins de reduzir o número de entradas e, assim, aumentar a relevância dos dados levados em consideração pela rede *Multilayer Perceptron* (MLP). A análise dos resultados obtidos comparou os resultados das previsões dos 3 modelos com os dados históricos e estudou seus erros de previsão por meio do MSE, RMSE, MAE, Desvio Padrão e teste abrangente de previsões.

Outra abordagem agora com algoritmos de *deep learning* é proposto por (HRASKO, 2015) com um algoritmo híbrido que combina uma GBRBM com FNN com a finalidade de que as unidades ocultas do RBM extraiam características latentes das séries temporais, na expectativa de se chegar a uma representação de alto nível, servindo como entrada para o algoritmo de aprendizado supervisionado FNN. Foram utilizados três *benchmarks* de séries temporais (Austrália Energy Produção, Conversão do Dólar para Libra e Oscilação do Atlântico Norte). Nos experimentos o modelo híbrido proposto GBRBM + FNN superou o algoritmo *Backpropagation* (BP) em 2 de 3 *benchmarks* em relação ao erro quadrático médio (EQM) como uma medida do erro de previsão.

Já em (LIU, 2015) utilizou-se uma rede neural convolucional de aprendizagem profunda (CNN) para a previsão de eventos de curto e longo prazo de movimentos de preços de ações do mercado de financeiro. Foram extraídos mais de 10 milhões de eventos de texto de notícias da plataforma *Bloomberg* do período de outubro de 2006 a novembro de 2013 e coletados os dados de índices e preços do *Yahoo Finance* dos códigos de ativos do S&P500. Os resultados em relação a medidas de desempenho acurácia e coeficiente de correlação

de *Matthews* mostram que o modelo proposto alcançou melhorias de quase 6% na previsão do índice S&P500 e previsão individual de ações respectivamente, em comparação ao *buy and hold* do período avaliado.

(JIA, 2016) propôs um modelo *Long Short-Term Memory* (LSTM) com utilização de dados diários (preços abertura, máximo, minimo e fechamento e volume) obtidos através do *Yahoo Finance* para o período de janeiro de 2005 a dezembro de 2015. Os dados coletados foram normalizados através de percentuais de retornos. O modelo utilizou 9 anos de dados para a fase de treinamento reduzindo, esta série para duas janelas deslizantes, realizando o treinamento de 1 época. Em seguida os dados são reduzidos para 4 janelas deslizantes, sendo repetido até o tamanho de 256 totalizando 100 épocas.

Já (XIONG, 2016) utilizou um modelo *Recurrent Neural Network* (RNN) com LSTM em bloco sem o emprego de camadas ocultas e em paralelo realizando outro treinamento através de 32 exemplos com o MAPE, 2 modelos de Regressão Linear (RIDGE e LASSO) e um modelo Autoregressivo (GARCH). Coletou dados diários (abertura, máximo, minimo, fechamento) do indicador S&P500 atraves do LSTM acessando o *Google Domestic Trends*, com 25 indicadores domésticos de tendencias como entrada para treinamento da rede. O período de treinamento foi de 19/10/2004 a 09/04/2012, totalizando 70% dos dados avaliados, e o período de teste foi de 12/04/2012 a 24/07/2015, totalizando 30%.

Em sua pesquisa (GIACOMEL, 2016) propôs um método de compra e venda de ações baseado nas previsões feitas por 2 *ensembles* de redes neurais adaptados para 2 perfis de investimento: investidores moderados e investidores agressivos. Os resultados desses ensembles preveem se determinada ação irá subir ou descer no próximo período ao invés de prever seus valores futuros, permitindo que se criem recomendações de operações de compra ou venda para o próximo período de tempo. Para mostrar a eficiência do nosso método em diferentes situações, o mesmo é avaliado exaustivamente em dois conjuntos de dados diferentes: os mercados de ações norte americano (S&P 500) e brasileiro (B3). Operações reais foram simuladas nestes mercados e fomos capazes de lucrar em 89% dos casos avaliados, superando os resultados das abordagens comparativas na grande maioria dos casos.

Considerando custos de corretagem, (NAMETALA, 2017) apresenta uma estratégia de negociação automatizada (robô investidor) que combina previsões feitas por redes neurais artificiais e preditores econométricos por meio de seletores inteligentes e comitês de decisão diversos que, progressivamente no tempo, são otimizados, gerando cenários lucrativos. Foi desenvolvida uma metodologia para seleção de ações que possuam histórico de alta liquidez e, aliado a isso, um algoritmo negociador que possui gestão de risco também otimizada temporalmente. Todos os dados das séries que serviram de base à pesquisa são reais, contudo o processo de negociação foi feito com capital simulado. Nos experimentos

realizados o robô investidor utilizou um histórico de 3 anos entre 2012 e 2013 para gerar estratégias de investimento aplicadas a dez ações de companhias nacionais e ao índice Ibovespa simultaneamente.

Para avaliar uma RBM na realização de seleção e extração de características dos dados, (LIU, 2017) incorporou uma RBM a vários classificadores para prever a tendência do mercado de ações a curto prazo, onde 11 indicadores técnicos foram inicialmente inferidos usando-se dados de negociação, por exemplo, preço abertura, preço mínimo, preço fechamento e preço máximo. Posteriormente, esses indicadores técnicos são transmitidos para valores usando uma camada de preparação determinística de tendência, que são aplicados a uma RBM para extrair características latentes e transformá-los em representações binárias para utilização como entrada de dados em outros modelos de classificação tradicionais. Nos experimentos compararam os desempenhos de 3 classificadores diferentes com base na entrada pré-processada por 4 abordagens, para 3 códigos de ações e em diferentes períodos.

Já o modelo proposto em (LEE, 2017) tem a finalidade de prever a receita futura de uma empresa, lucro operacional e lucro líquido usando dados financeiros e de patentes passados nos períodos de 2000 e 2015. Inicialmente inclui uma fase de aprendizado não supervisionada, usando um RBM com todo o conjunto de dados de treinamento realizando a extração e seleção de características latentes. Em seguida, com estes dados de treinamento transformados, ocorre uma fase de ajuste fino através do algoritmo *Backpropagation*. Esses dados refletem as últimas tendências na correlação entre os preditores e desempenho corporativo na previsão, a fim de melhorar a precisão da previsão da rede, utilizando como métrica de avaliação de desempenho, o valor do erro quadrático médio (RMSE).

Em outro trabalho (KIMURA, 2017) foram testados modelos de aprendizado de máquina (SVM, *Bagging*, *Boosting*, e *Random Forest*) para prever a falência de empresas um ano antes do evento e seu desempenho é comparado com os resultados de análise discriminante, regressão logística e redes neurais. Utilizaram-se dados de 1985 a 2013 sobre empresas norte-americanas, integrando informações do banco de dados *Salomon Center* e *Compustat*, analisando mais de 10.000 observações. Comparando os melhores modelos, com todas as variáveis preditivas, a técnica de aprendizado de máquina relacionada à *Random Forest* levou a 87% de precisão, enquanto a Regressão Logística e análise discriminante linear levaram a 69% e 50% de precisão, respectivamente.

Em (BAUGHMAN, 2018) foi utilizada uma rede neural do tipo *Long-Short Term Memory (LSTM)* para previsão do preço à vista das instâncias $spot$ do *Amazon EC2*. Estas instâncias são a capacidade computacional extra na Nuvem AWS, disponíveis com grandes descontos em relação ao preço das instâncias sob demanda. Configurou-se o modelo LSTM proposto variando o número de camadas na rede entre 1 e 4 e a largura de cada camada entre 16 e

128 nós. Por fim, comparou-se uma rede simples de 3 camadas composta de duas camadas LSTM (cada uma com 32 unidades) com o modelo de linha de base ARIMA usando o ponto histórico dos dados de preços, obtendo-se uma redução de erro de treinamento em até 95%.

(ASSIS, 2019) desenvolveu 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. Realizou-se testes com os ativos BOVA11, CIEL3, ITUB4, PETR4, USIM5, CMIG4, GGBR4, KROT3 e GOLL4, e foram encontrados uma classificação com até 57%, além de resultados financeiros com ganhos de até 100% do valor de capital inicialmente investido.

Já (MIRANDA, 2019) construiu um simulador de bolsa de valores, um conjunto de indicadores baseados em microestrutura de mercado, uma técnica de rotulação de séries de preço e, por fim, um agente autônomo de negociação. A base de dados escolhida para a validação experimental consiste dos dados de negociação de 2018 do minicontrato futuro de dólar da B3. Os resultados financeiros obtidos pelo agente autônomo são avaliados em diferentes cenários considerando-se custos de operação e latências de rede.

Devido à grande quantidade de técnicas e modelos existentes na literatura para a realização de previsão de séries temporais, os critérios de seleção definidos na revisão sistemática da literatura, foram direcionados esforços em contemplar somente trabalhos relacionados ao escopo de algoritmos de aprendizado de máquina tradicionais de categoria supervisionados, e combinações entre algoritmos de aprendizagem profunda (*deep learning*) contendo atuação como classificadores de previsão de tendências para o mercado de ações.

A Tabela 3 apresenta de forma consolidada os trabalhos relacionados que foram selecionados, destacando os seguintes pontos: referências, fontes de dados, algoritmos / modelos de previsão de tendência e, por último, as medidas de avaliação e desempenho (Métricas) apresentadas como resultados. Foram selecionados 23 trabalhos, sendo 11 artigos ((LIU, 2017) (LEE, 2017) (KIMURA, 2017) (KLABJAN, 2016) (XIONG, 2016) (JIA, 2016) (HRASKO, 2015) (LIU, 2015) (CAO, 2015) (SHARANG, 2015)), 11 dissertações de mestrado ((NAMETALA, 2017) (FERREIRA, 2016) (SANTOS, 2016) (MANSOURI, 2016) (ESTRADA, 2015) (AAMODT, 2015) (ZHANG, 2015) (ROCA, 2015) (SILVA, 2014) (RAMALHO, 2014) (VENGERTSEV, 2014)) e 2 teses de doutorado ((BOOTH, 2016) e (ASSIS, 2019)). De forma convergente, realizam a utilização de indicadores financeiros, análise de informações da Web, técnicas de ciência dos dados (normalização e transformação, caracterização e análise) e aplicações de aprendizado de máquina.

Os algoritmos mais utilizados de forma independente e ou combinada foram Máquina

Tabela 3 – Comparação entre os trabalhos relacionados selecionados pela RSL.

Referências	Fontes de Dados	Algorítmos e Modelos	Arquiteturas	Métricas e Resultados
(MIRANDA, 2019)	B3	Indicadores	Python	Acurácia de 41,23%
(ASSIS, 2019)	B3	Metaclassificadores	Java e Python	Acurácia de 57%
(GHOSH et al., 2019)	B3	LSTM	Python	Acurácia de 49,38%
(BAUGHMAN, 2018)	B3	LSTM	Não informado	Acurácia de 63%
(LIU, 2017)	Índice FTSE 100	RBM e SVM	Não informado	MAPE de 95%
(LEE, 2017)	Dados corporativos	RBM + BP e DBN	Java	MAPE de 3880,24 e 4288,95
(KIMURA, 2017)	Dados corporativos	SVM e Random Forest	Não informado	Acurácia de 50%
(NAMETALA, 2017)	B3	RNA	Python	Acurácia de 53% a 61,4%
(KLABJAN, 2016)	B3	DNN	Python, C++	Acurácia de 42%
(XIONG, 2016)	Índice S\&P500	LSTM e RIDGE	Python	MAPE de 31%
(JIA, 2016)	Índice S\&P500	LSTM, RIDGE e LASSO	Python	RMSE de 26,50%
(BOOTH, 2016)	Índices FTSE100	MLP, SVR e RF	Matlad	Acurácia de 71,57% a 75,74%
(FERREIRA, 2016)	StockTwits	N/A	Crisp-DM	Correlação > 0,5
(SANTOS, 2016)	Twitter, Bovespa	N/A	Não informado	estimativas de uso
(MANSOURI, 2016)	Índice S\&P100	LSTM	Não informado	Acurácia de 77,00%
(HRASKO, 2015)	Dow Jones	RBM+FNN e BP	Não informado	MAPE de 31,54%
(LIU, 2015)	Índice S\&P500	WB-NN e WB-CNN	Não informado	Acurácia de 64,21% e 65,48%
(CAO, 2015)	Dow Jones	ARIMA, RL e ANN	Não informado	Acurácia de 67,34%
(SHARANG, 2015)	US Treasury	PCA, DBN e RBM	Python	Acurácia de 60,00%
(ESTRADA, 2015)	Índice S\&P500	RBM e DBN-MLP	Python, Matlab, C++	Acurácia de 50,84% e 47,11%
(AAMODT, 2015)	S\&P500, NASDAQ	RBM+SVR e CNN	Python	Acurácia de 50,93% e 50,74%
(ZHANG, 2015)	Dow Jones	RBM, DBN RNN, e SVM	Não informado	Precisisão de 76%
(ROCA, 2015)	B3	PCA, MLP e GARCH	NeuroShell, Brainmaker	estimativas de uso
(SILVA, 2014)	B3	NB, RL e SVM	SQL Server 2012 Express	Acurácia de 61%
(RAMALHO, 2014)	B3	MLP + BP	Matlab e outros	Acurácia de 55%
(VENGERTSEV, 2014)	B3	RBM + MLP, ARIMA e SVR	Matlab e outros	Acurácia de 52%

Restrita de *Boltzmann (RBM)* com 31% e, em seguida as *Long Short-Term Memory* (LSTM) com 19%. Desta forma, visualiza-se a possibilidade de utilização do algoritmo RBM de categoria não-supervisionada associada a outros algoritmos de classificação tradicionais de categoria supervisionada, que serão detalhados no Capítulo 3 de fundamentação teórica.

Já com relação às métricas de avaliação empregadas nos algoritmos e modelos de previsão de tendências utilizados nos trabalhos selecionados, cerca de 80% utilizam Acurácia e Precisão e o restante 20% estão distribuídas entre as medidas estatísticas de estimação de erro, como, por exemplo, MSE, RMSE e MAPE.

Devido ao tema desta tese ser interdisciplinar, o principal desafio comum a todos os trabalhos relacionados, Tabela 3, é tentar caracterizar, compreender e predizer o comportamento dos preços e consequentemente a direção do fluxo das tendências dos ativos negociados no mercado de ações, independente da técnica utilizada.

Com o propósito de conhecer o estado da arte atual para propor direções e novos caminhos a serem investigados, verificaram-se várias dificuldades em replicar os experimentos realizados nestes trabalhos relacionados, devido a não existir um padrão de metodologias, arquiteturas, ferramentas, acesso às fontes de dados, dentre outros, comum aos trabalhos relacionados.

Após o levantamento dos trabalhos relacionados, observou-se, a necessidade de automatização da geração de gráficos e tabelas e de padronizar a utilização de algoritmos / modelos

de previsão de tendências em uma única arquitetura.

Com a finalidade de implementar modelos classificadores de previsão de tendências, que pudessem ser parametrizáveis, e disponibilizado para a comunidade acadêmica, porpõe-se o desenvolvimento de um arcabouço automatizado para a modelagem e a avaliação de estratégias de investimento, utilizando, para isso, conceitos de teorias econômicas, dados de cotações históricas e algoritmos de aprendizado de máquina.

Desta forma, esta tese diferencia-se dos demais trabalhos relacionados da Tabela 3, devido ao arcabouço proposto, gerar de forma automatizada, através da Linguagem *Python*, gráficos e tabelas contendo os resultados estatísticos, desempenho (acurácia, precisão e retorno financeiro), risco, custos operacionais, dentre outros, para os 8 modelos classificadores de previsão de tendências que serão detalhados no Capítulo 3.

Capítulo 3

Fundamentação Teórica

Neste Capítulo serão apresentados alguns conceitos sobre o mercado de ações de valores, tais como: análise técnica, indicadores técnicos, estratégias de negociação e custos operacionais. Em seguida, serão introduzidos os algoritmos de aprendizado de máquina utilizados como modelos de previsão de tendências e algumas métricas de avaliação de desempenho e risco.

3.1 Introdução ao Mercado de Ações

De acordo com (PINHEIRO, 2019) o mercado de capitais surgiu como uma alternativa ao mercado de financiamento tradicional bancário de crédito, com a finalidade de intermediar negociações entre investidores que gostariam de investir em empresas e empresas que precisam de captar recursos para seus projetos, o que possibilita investidores encontrar oportunidades de investimentos através de empresas que estão dispostas a aceitar novos sócios.

O mercado de capitais é um segmento do sistema financeiro que disponibiliza e engloba investimentos de categorias de renda fixa e de renda variável, como por exemplos, os fundos de investimentos, títulos de propriedades (ações), títulos de dividas (debentures), operações estruturadas, letras de créditos imobiliários (LCI), letras de crédito do agronegócio (LCA), dentre outros.

Segundo (MARQUES, 2013) o mercado de capitais é um conjunto de operações de transferência de recursos financeiros de prazo médio, longo ou indefinido, efetuadas entre agentes poupadores e investidores, através de intermediários financeiros. No entanto, esta pesquisa abrange especificamente a categoria de renda variável, em especial o mercado de ações, no qual serão apresentados algumas características, particularidades, conceitos básicos e fundamentos.

O mercado de ações centraliza a negociação de diversos tipos de papéis ao redor do mundo, como ações e derivativos. Ele permite que investidores comprem e vendam esses papéis de acordo com suas preferências e necessidades. O preço desses papéis é definido, genericamente, pelas leis de oferta e demanda.

Ações (ativos) são títulos que representam frações de capital de uma determinada 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 para ser negociada na bolsa de valores (ISHIKAWA, 2000), sendo constituída de dois tipos de ações:

- **Ordinárias:** estas ações proporcionam a participação nos resultados da empresa e conferem o direito de voto nas assembleias da empresa; e
- **Preferenciais:** este tipo de ações o investidor terá preferência no recebimento de dividendos, no entanto, não terá o direito a voto nas assembleias.

O mercado de ações é constituído de bolsas de valores, corretoras e investidores, sejam eles empresas ou pessoas físicas. No Brasil, a Brasil Bolsa Balcão (B3) (B3, 2019a), resultado da fusão entre BMF&Bovespa e Cetip, é responsável pela negociação dos papéis de empresas brasileiras e de outros derivativos. A B3 é a companhia responsável pela administração de mercados organizados de títulos, valores mobiliários e contratos derivativos no Brasil, realizando serviços de registro, compensação, liquidação, divulgação, atuando como contraparte central garantidora da liquidação financeira das operações realizadas em seus ambientes de forma transparente (B3, 2014).

3.1.1 Análise Técnica

A análise técnica surgiu no século XX através das teorias propostas por *Charles Dow* (HAMILTON, 1922) e engloba tanto os dados numéricos relativos quanto os gráficos das empresas que mostrava o comportamento das ações ao longo do tempo. Alguns princípios em sua teoria que, apesar de mais de 100 anos de existência, são utilizados atualmente:

- **Os preços refletem tudo** e todas as opiniões sobre um determinado ativo, bem como todas as notícias públicas são levadas em consideração e estão expressas nos preços dos códigos de ativos;
- **O mercado segue tendências** categorizadas em alta, baixa ou neutra. Alta quando o resultado de uma força compradora, ou seja, os compradores estão dominando o mercado. Tendência de baixa o inverso da anterior. E tendência neutra, quando ocorre um equilíbrio entre comprador e vendedor;
- **O volume de negociações acompanha a tendência** em um mercado de alta, maior força compradora, o volume de negociações aumenta quando o preço do ativo au-

menta e o volume diminui quando o preço do ativo diminui. Em um mercado onde a força vendedora é superior à força compradora, ou seja, um mercado de baixa, o volume tem uma relação inversa com o preço, isto é, o volume de negociações aumenta quando o preço do ativo diminui e o volume diminui quando o preço do ativo aumenta;

- **Uma tendência é finita** e sua confirmação pode ser verificada com a ajuda de gráficos e índices de preços simultaneamente;
- **Retas de Suporte** é um patamar de preços a partir do qual grande parte dos investidores passa a sentir segurança para aplicar num papel e, com isso, o volume de negócios intensifica-se; e
- **Retas de Resistência** é um teto ou limite de preços momentâneo, a partir do qual uma parte dos investidores resolve vender suas ações por achar que o preço já alcançou o seu valor máximo, determinando o aparecimento de certa quantidade de ordens de venda naquele valor.

Para os analistas técnicos, as diversas variáveis fundamentais, políticas, psicológicas ou de qualquer outra ordem estão refletidas nos preços de mercado, por isso dizer que os preços refletem todos os fatores envolvidos no mercado. O propósito de representar a ação dos preços em um gráfico é identificar tendências.

Outro recurso utilizado na análise técnica são os indicadores técnicos, os quais utilizam uma série de dados que são derivados pela aplicação de uma fórmula para os dados referentes a um título, sendo classificados como:

- 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*); e
- Indicadores de Volume: são baseados no fato de que o volume precede o movimento do preço (ex.: OBV, Oscilador de Volume, *Chaikin Oscilador*).

Os dados mais comuns utilizados em indicadores são preços (abertura, máxima, mínima e fechamento) e ainda podem conter informações de volume ou contratos em aberto em suas fórmulas.

3.1.2 Estratégias de negociação

De acordo com (B3, 2019b), os tipos de estratégias de negociação para investimentos em ações na bolsa de valores são caracterizados em relação à janela de tempo de duração entre o momento da compra e venda de um determinado ativo, descritos a seguir:

- ***Day trade*** ou ***Intraday***: realizar a compra e venda no mesmo dia, ou seja, quando a operação é concretizada inteiramente no intervalo de tempo de um pregão;
- ***Swing trade***: realizar a compra e venda de ativos após o fechamento de um pregão e no máximo em 10 dias; e
- ***Position trade***: realizar a compra e venda de ativos em um intervalo que pode durar algumas semanas, meses ou anos.

Com o desenvolvimento das plataformas de negociação, atualmente existem duas formas de realizar estratégias de negociação:

- **Operação manual**: no qual o investidor desenvolve estratégias próprias, necessitando de sua monitoria constante e sua atuação para a realização de ordens de compra e venda de ativos via a plataforma de operação *Home Broker* da sua corretora; e
- **Operação por algoritmos (*algotrading*)**: é um software que utiliza um sistema de cálculos baseados em modelagem matemática, seguindo uma estratégia rigorosa, que pode trabalhar 24 horas por dia, 7 dias por semana, por um tempo ininterrupto em vários mercados. A execução destes *algotrading* depende de integração e compatibilidade com a plataforma de operação *Home Broker* das corretoras.

O *algotrading* pode ser utilizado também de forma híbrida em complemento a uma análise humana que avalia situações de mercado sendo complementado com análise fundamentalista e atuando como um indicador de previsão de tendências para auxiliar na tomada de decisão.

3.1.3 Custos Operacionais

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 que 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;

- **Emolumentos e liquidação**: os emolumentos são cobrados pela Brasil Bolsa Balcão (B3), 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, tem 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; e
- **Imposto de Renda**: o lucro obtido ao se operar renda variável, o chamado ganho líquido, está sujeito à cobrança de imposto de renda. Ganhos líquidos com operações *daytrade* (compra e venda de um ativo no mesmo dia) são tributados em 20%.

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 os diversos mercados de investimento estão disponíveis em (B3-VARIáVEL, 2020), (B3-AçõES, 2020) e (B3-CUSTóDIA, 2020).

3.2 Algoritmos de Aprendizado de Máquina

Aprendizado de máquina são métodos de análise de dados que automatizam o desenvolvimento de modelos analíticos, usando algoritmos que aprendem interativamente a partir de dados, permitindo que os computadores encontrem respostas e ou representações ocultas sem serem explicitamente programados para procurar algo específico (SHALEV-SHWARTZ; BEN-DAVID, 2014). A seguir são definidos os 4 principais grupos de modelos de aprendizado de máquina:

- **Supervisionado:** é apresentado ao modelo exemplos de entradas e saídas desejadas, com a finalidade de aprender uma regra geral, que mapeia as entradas para as saídas. Na fase de treinamento estes algoritmos avaliam o conjunto de entrada e saída de dados conhecidos e, em seguida, nas fases de validação e teste, através de padrões aprendidos na fase anterior, recebem uma série de entrada e tenta realizar a previsão da sua respectiva saída.
- **Não supervisionado:** esses algoritmos são utilizados em dados que não possuem rótulos históricos, no qual o sistema não sabe a "resposta certa". O objetivo é descobrir o que está sendo mostrado com a exploração dos dados e encontrar alguma estrutura

e ou representação sendo melhor empregado em dados transacionais;

- **Semi-supervisionado:** esses algoritmos são utilizados para as mesmas aplicações que o aprendizado supervisionado, porém eles utilizam tanto dados rotulados quanto não marcados para o treinamento. Normalmente uma pequena quantidade de dados rotulados com uma grande quantidade de dados não rotulados, pois os dados não rotulados são mais baratos e precisam de menos esforço para serem adquiridos; e
- **Por reforço:** esses algoritmos são comumente utilizados na robótica, jogos e navegação. Com o aprendizado por reforço, o algoritmo descobre pela tentativa e erro quais ações geram as maiores recompensas. Este tipo de aprendizagem tem três componentes principais: o agente (o aluno ou tomador de decisões), o ambiente (tudo com o qual o agente interage) e as ações (o que o agente pode fazer).

Nas técnicas tradicionais de aprendizagem de máquina, a qualidade dos algoritmos depende muito da representação dos dados em certas características e o pré-processamento consome assim grande parte do esforço do analista e é específico para cada domínio (BENGIO, 2009).

Devido ao grande volume de dados não estruturados disponíveis em vários formatos, aplicações, plataformas e principalmente na web, surge uma necessidade de aprendizagem em grandes quantidades de dados de uma forma não-supervisionada (dados não anotados), com a realização de abstrações complexas dos dados através de um processo de aprendizagem hierárquica muito similar ao que ocorre no cérebro humano, conhecida como aprendizagem profunda (*deep learning*).

Diante dos trabalhos selecionados pela RSL na Tabela 3 do Capítulo 2, serão avaliados nesta pesquisa os algoritmos tradicionais de aprendizado supervisionado Árvore de Decisão *(AD J48)*, *Naive Bayes* (NB), Máquina de Suporte de Vetores *(SVM)* e *K-Nearest Neighbor (KNN)* e, o algoritmo de aprendizado não supervisionado de *deep learning* Máquina Restrita de *Boltzmann (RBM)*.

3.2.1 Árvore de Decisão (AD J48)

AD J48 é um algoritmo que utiliza o método divisão e conquista para aumentar a capacidade de predição das árvores de decisão. Assim, sempre utiliza o melhor passo avaliado localmente, sem se preocupar se esse passo vai produzir a melhor solução. Desta forma, um problema é dividido em vários subproblemas sendo criadas sub-árvores entre a raiz e as folhas. O resultado final é uma árvore com nós de decisão e nós de folha. O nó superior de decisão corresponde ao melhor preditor denominado de nó raiz que pode manipular dados tanto categóricos quanto numéricos (QUINLAN, 1993).

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 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. A classificação de instâncias parte 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.

3.2.2 *Naive Bayes* (NB)

O NB (PEARL, 1988) é um algoritmo que trabalha com a representação do conhecimento incerto e incompleto por meio do Teorema de *Bayes*, com os pressupostos de independência entre os preditores. Uma NB, também chamada de rede probabilística ou rede causal, pode ser vista como um modelo que utiliza teoria dos grafos, condições de *Markov* e distribuição de probabilidades para representar uma situação, suas variáveis e estados e então realizar inferências.

Os grafos direcionados acíclicos em NB são chamados de estrutura, e os valores da distribuição de probabilidades condicionais são chamados de parâmetros. Tanto a estrutura quanto os parâmetros de uma NB 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. 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. Um importante aspecto de uma NB 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.

3.2.3 Máquinas de Suporte de Vetores (SVM)

As Máquinas de Vetores de Suporte (SVM) são embasadas pela teoria de aprendizado estatístico, desenvolvida por Vapnik (VAPNIK, 1995), na qual 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 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 positivos e os pontos negativos mais próximos do hiperplano, conhecidos como suportes de vetores. 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 algorítimos 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. Alguns dos kernels mais utilizados na prática são os Polinomiais, os Gaussianos ou RBF (Radial-Basis Function) e os Sigmoidais.

3.2.4 *K-Nearest Neighbor (KNN)*

O *KNN* é um algoritmo usado para classificação e regressão, formado por vetores n-dimensionais e cada elemento deste conjunto representa um ponto no espaço n-dimensional, que tenta determinar a classe de um elemento que não pertença ao conjunto de treinamento (RAMAKRISHNAN, 1999).

A finalidade é procurar K elementos do conjunto de treinamento que estejam mais próximos deste elemento desconhecido, ou seja, que tenham a menor distância, verificando quais são as classes desses K vizinhos e a classe mais frequente será atribuída à classe do elemento desconhecido. A distância pode, em geral, ser qualquer medida métrica: a distância euclidiana padrão é a escolha mais comum. Os métodos baseados em vizinhos são conhecidos como métodos de aprendizado de máquina não generalizados, uma vez que já é conhecido todos os seus dados de treinamento transformados em uma estrutura de indexação rápida.

3.2.5 Máquina Restrita de *Boltzmann (RBM)*

As *RBM* (SALAKHUTDINOV, 2006) são redes neurais de aprendizado não-supervisionado. Estas são caracterizadas principalmente por sua capacidade de aprenderem representações internas e de resolverem problemas combinatórios complexos. A camada visível corresponde aos componentes de um exemplo de entrada como, por exemplo, os *pixels* de uma imagem. A camada escondida modela a dependência entre os componentes da camada de entrada, que por sua vez deverá aprender a extrair características desses dados.

As conexões entre neurônios são bidirecionais e simétricas, apresentando um tráfego de

informação em ambos os sentidos da rede, com conexões entre neurônios de camadas diferentes, justificando portanto a denominação máquina restrita. Em sua topologia existem quatro hiper-parâmetros: quantidade de neurônios da camada visível (v), quantidade de neurônios da camada escondida (h), a taxa de aprendizado (lr) e a quantidade de ciclos (ep). Uma taxa de aprendizado muito baixa torna o aprendizado da rede muito lento, ao passo que uma taxa de aprendizado muito alta provoca oscilações no treinamento e impede a convergência do processo de aprendizado. Geralmente seu valor varia de $0,1$ a $1,0$ (SIMON, 1998). Já o número de ciclos é o número de vezes em que o conjunto de treinamento é apresentado à rede. Um número excessivo de ciclos pode levar a rede à perda do poder de generalização (*overfitting*). Por outro lado, com um pequeno número de ciclos, a rede pode não ser capaz de modelar o comportamento geral do sistema (*underfitting*).

3.3 Métricas de Avaliação

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

As métricas de desempenho tem objetivo de realizar a avaliação da qualidade dos classificadores (preditores) baseados em modelos de aprendizado de máquina, sendo constituídas de fórmulas matemáticas e estatísticas (ZHANG, 2015). As métricas mais utilizadas são a Acurácia, Revocação, Precisão, *F1-score* e Especificidade assim detalhadas:

- **Acurácia**: é a quantidade de amostras positivas (AP) e negativas (AN) classificadas corretamente dividido pelo total de amostras (TA) da série avaliada, representada pela Equação 1.

$$Acurácia = \frac{AP + AN}{TA} \tag{1}$$

- **Precisão**: é a quantidade de amostras positivas (AP) classificadas corretamente sobre o total de amostras classificadas como falsas positivas (FP) acrescida das amostras positivas em percentual, representada pela Equação 2.

$$Precisão = \frac{AP}{FP + AP} \tag{2}$$

- ***Revocação***: é a quantidade de amostras positivas (AP) classificadas corretamente sobre o total de amostras classificadas como falsas negativas (FN) mais AP em percentual, representada pela Equação 3.

$$Revocação = \frac{AP}{FN + AP} \tag{3}$$

- ***F1-score***: é a média harmônica entre a Precisão e o *Revocação*, representada pela Equação 4.

$$\textit{F1-score} = \frac{2 * Precisão * \textit{Revocação}}{Precisão + \textit{Revocação}} \tag{4}$$

- **Especificidade**: é a quantidade de amostras negativas identificadas corretamente (AN) sobre o total de amostras negativas (TAN), representada pela Equação 5.

$$Especificidade = \frac{AN}{TAN} \tag{5}$$

As métricas de medidas de risco tem a finalidade de avaliar o risco versus retorno (ERNST, 2014) (VARGA, 2001) das estratégias de negociações proposta utilizando os seguintes indicadores: volatilidade, índice *Sharpe* e índice *Sortino*.

- **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, onde Pi é o preço de fechamento no período i e $Pi - 1$ é o mesmo preço de fechamento no período anterior e ln é o logaritmo natural, conforme Equações 6 e 7. O valor 252 é o período anualizado em dias úteis.

$$r = ln(\frac{Pi}{Pi - 1}) \tag{6}$$

$$Volatilidade = (\sqrt{\frac{\sum_{|r-\bar{r}|}{}^{2}}{n}}) * \sqrt{252} \tag{7}$$

- **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 do Ibovespa do período avaliado e o resultado dividido pela volatilidade conforme Equação 8

$$Sharpe = (RF - Ibovespa)/volatilidade \tag{8}$$

- **Índice *Sortino***: é uma adaptação do índice de *Sharpe*, na qual a volatilidade do ativo é substituída pela *volatilidadeDownside*, ou seja, utilizando apenas os retornos abaixo de um certo valor de referência nas Equações 6 e 7. A vantagem do índice de *Sortino* com relação ao de *Sharpe* é que ele reflete apenas a volatilidade "ruim", ou seja, das perdas, conforme Equação 9

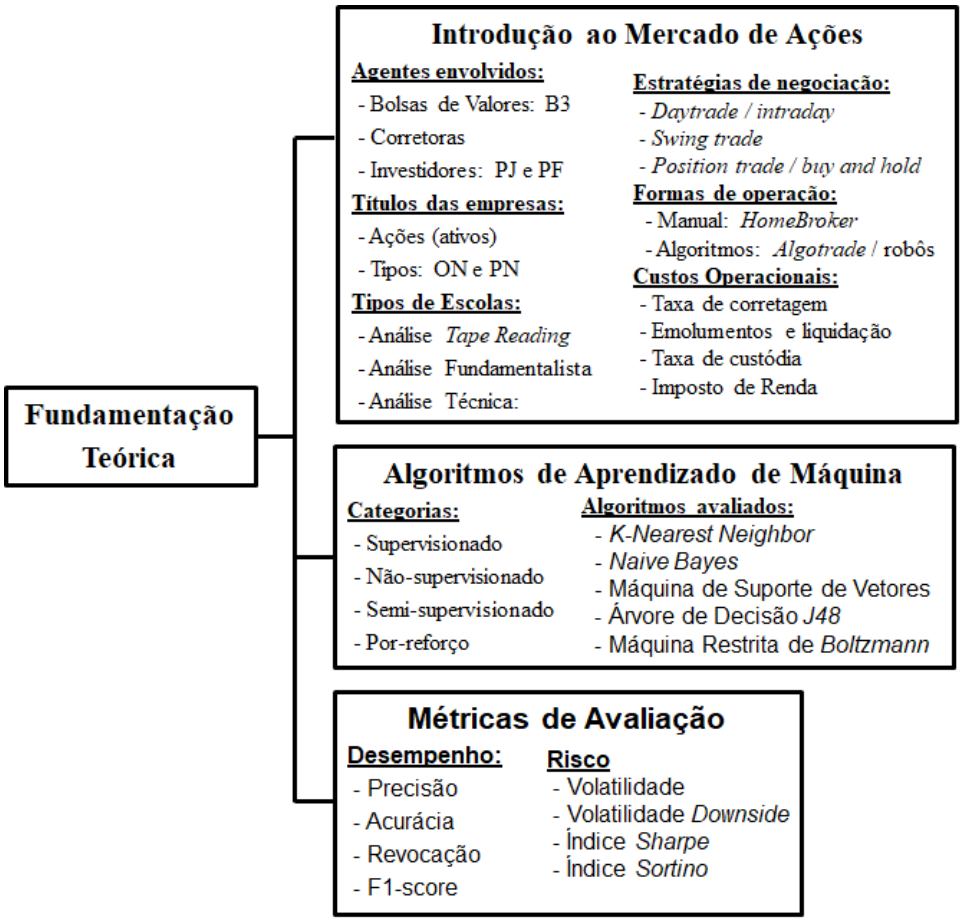
$$Sortino = (RF - Ibovespa)/volatilidadeDownside \tag{9}$$

Optou-se por utilizar como taxa de referência do mercado de renda variável nas Equações 7 e 9 o índice Ibovespa, sendo um indicador de desempenho das ações mais negociadas na Bolsa Brasil Balcão (B3).

3.4 Considerações Finais

Na Figura 2 é apresentado um resumo dos 3 principais tópicos que compõe a fundamentação teórica: Introdução ao Mercado de Ações, Algoritmos de Aprendizado de Máquina e Métricas de Avaliação.

Figura 2 – Resumo da Fundamentação Teórica utilizada: Mercado de Ações, Aprendizado de Máquina e Métricas de Avaliação.



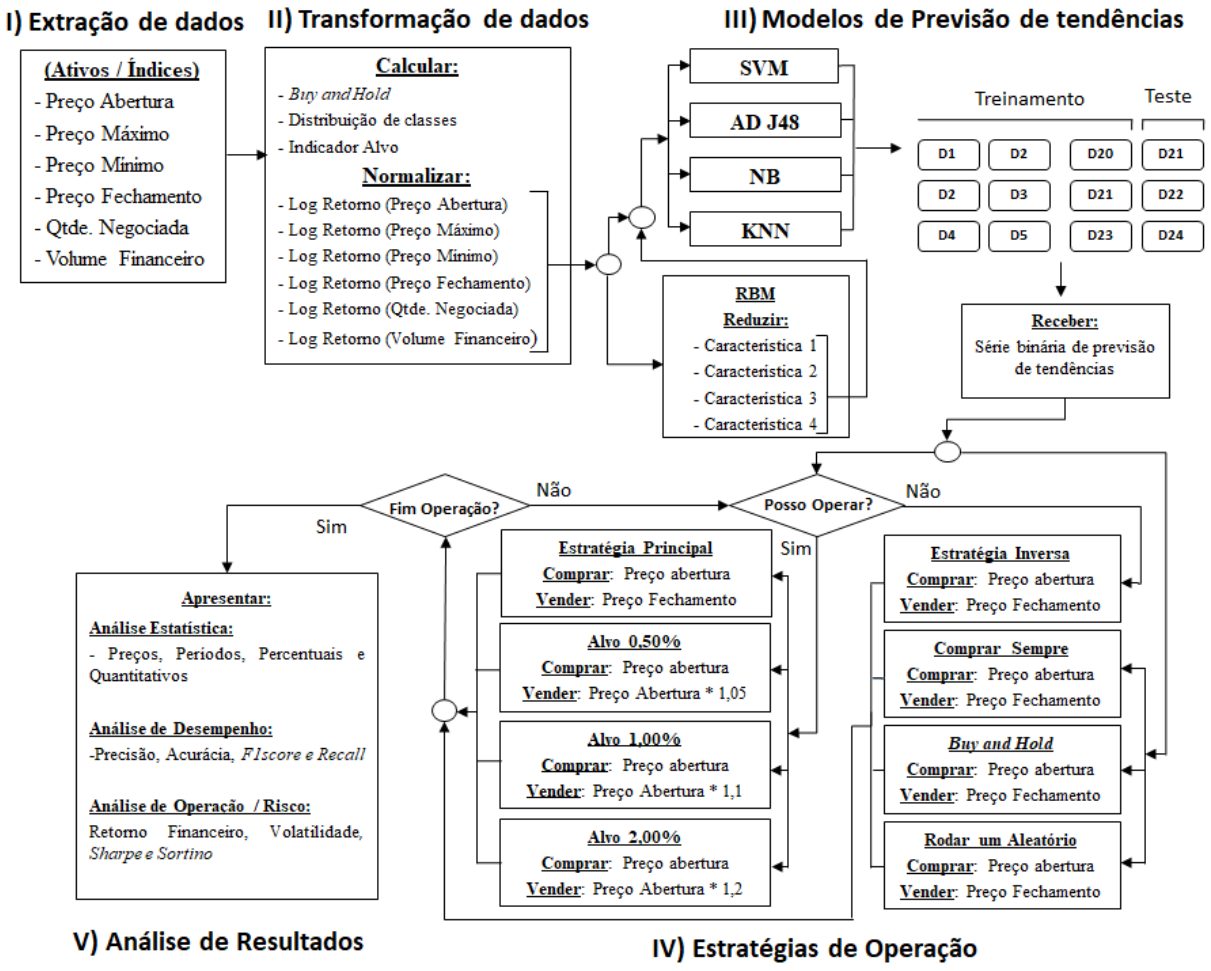
Realizou-se neste Capítulo a descrição de alguns conceitos, técnicas e fundamentos teóricos sobre o mercado de ações, métodos de análises, estratégias de negociação mais utilizadas, algoritmos de aprendizado de máquina e as métricas de avaliação de desempenho de medidas de risco das estratégias de negociação, contextualizando com o estado da arte em relação aos principais trabalhos selecionados pela revisão sistemática da literatura realizada no Capítulo 2.

Capítulo 4

Metodologia

Este Capítulo descreve o arcabouço proposto implementado, constituído de 5 etapas: (I) Extração de Dados, (II) Caracterização e Transformação de Dados, (III) Modelos de Previsão de Tendências, (IV) Estratégias de Operação e (V) Análise de Resultados, conforme detalhado na Figura 3.

Figura 3 – Arcabouço proposto.



4.1 Extração de Dados

Para desenvolver a presente pesquisa ativos listados na Brasil Bolsa Balcão (B3), foram coletados por meio do terminal *Bloomberg*[1] que disponibiliza informações, dados de preço, indicadores e notícias do mercado financeiro e de capitais. Esta plataforma disponibiliza o serviço *Bloomberg Professional* conectando mais de 320.000 profissionais de negócios, finanças e entidades governamentais a uma rede global e dinâmica de informações, notícias, pessoas e ideias, viabilizando decisões mais rápidas e eficazes. O acesso a esta plataforma foi disponibilizado pelo Laboratório do Grupo de Finanças (GFin) do Centro Federal de Educação Tecnológica de Minas Gerais Campus VI[2].

Realizou-se a extração destes dados de forma manual via *download* de arquivos de cotações históricas em formato (.csv) através da própria plataforma que disponibiliza um função parametrizável que agrupa a série temporal solicitada por granularidade (diária, semanal, mensal, etc.) juntamente com seus respectivos atributos relativos ao código do ativo, preços, dentre outros.

Os dados coletados são séries históricas dos preços (fechamento, abertura, máximo e mínimo), já ajustados, além da quantidade de negócios, volume financeiro e data (ano, mês e dia) das transações realizadas dos códigos e ativos avaliados na granularidade diária, após um determinado critério de seleção.

4.2 Transformação de Dados

Após o critério de seleção dos dados de cotações históricas é realizada a caracterização por setores de mercado, o cálculo do *buy and hold* de forma separada dos períodos de treinamento e de validação (teste) e, por último, é realizado o cálculo e apresentação das distribuições de classe 0 (alta) e de classe 1 (baixa).

Tabela 4 – Exemplo de modelo original da série histórica de preços.

Data	Preço de Abertura	Preço Máximo	Preço Mínimo	Preço Fecha.	Quantidade de papeis	Volume Fiannceiro
20150105	R$ 22,58	R$ 22,70	R$ 21,90	R$ 22,18	2258	12125198900
20150106	R$ 22,23	R$ 22,77	R$ 21,98	R$ 22,49	2223	16282625300
20150107	R$ 22,94	R$ 24,08	R$ 22,74	R$ 23,48	2294	9537481900
20150108	R$ 23,77	R$ 23,98	R$ 23,22	R$ 23,56	2377	7083603700
20150109	R$ 23,50	R$ 23,55	R$ 22,51	R$ 22,54	2350	8378459200
20150112	R$ 22,56	R$ 23,00	R$ 22,02	R$ 22,15	2256	7348127500

[1] *Bloomberg: www.bloomberg.com.br*
[2] GFIN: *www.dcsa.cefetmg.br/pesquisa/grupos-cnpq/gfin*

Na Tabela 4 apresenta-se exemplos dos formatos originais dos 6 atributos de preços (fechamento, abertura, máximo e mínimo), quantidade de negócios e volume financeiro, de granularidade diária, que são extraídos dos dados de cotações históricas de ativos da B3 ou de índices globais.

A normalização destes 6 atributos são realizadas aplicando o retorno de log natural (X), através da diferença entre preço inicial (Pi) e o preço anterior $(Pi-1)$, conforme a Equação 10.

$$X = log_{(Pi)} - log_{(Pi-1)} \tag{10}$$

Para retirar os ruídos (*outliers*) da série de retorno de log natural (X) e, consequentemente, facilitar o trabalho dos modelos de previsão, agrupou-se esta mesma série temporal (X) em α partes para que estas várias amostras de (μ) e (σ) possam representar a série em questão avaliada, sendo detalhado o cálculo na Equação 11.

$$\frac{\chi - (\sum_{j=i}^{\alpha} \mu)}{\sum_{j=i}^{\alpha} \sigma} \tag{11}$$

Quando uma série temporal de preço (X) é muito grande, a sua média (μ) e o seu desvio padrão (σ) podem não representar a realidade. Realizou-se simulações variando o α entre [5;15], no qual o valor 7 foi o mais adequado às séries temporais avaliadas.

Tabela 5 – Exemplo de modelo normalizado da série histórica de preços.

Data	Preço Abertura	Preço Máximo	Preço Mínimo	Preço Fecha.	Quantidade de papeis	Volume Fiannceiro	Classe Y (alvo)
20150105	0,2870	0,5010	0,6665	0,5874	0,5528	0,5673	1
20150106	0,4066	0,2711	0,0181	0,0893	-0,1810	-0,1632	1
20150107	-0,1738	0,0798	-0,0230	0,0107	0,4301	0,4393	0
20150108	0,2870	0,5010	0,6665	0,5874	0,5528	0,5673	0
20150109	0,4066	0,2711	0,0181	0,0893	-0,1810	-0,1632	0
20150112	-0,1738	0,0798	-0,0230	0,0107	0,4301	0,4393	0

O atributo indicador alvo *ClasseY* é calculado através da diferença entre os preços de Fechamento e Abertura do próximo dia $(D+1)$ da série temporal avaliada. Caso este resultado seja maior do que 0, o atributo alvo *ClasseY* recebe o valor de "1". Caso contrário, este atributo alvo recebe o valor de 0.

Como resultado da normalização realizada dos 6 atributos acrescidos do atributo indicador alvo *ClasseY*, são gerados 7 novos atributos de granularidade diária, que serão utilizados

como dados de entrada dos modelos de classificação que serão propostos, conforme o formato no exemplo da Tabela 5.

4.3 Modelos de Previsão de Tendência

Para a etapa de classificação, inicialmente utiliza-se 4 algoritmos de aprendizado de máquina tradicionais da literatura (*SVM, AD J8, NB* e *KNN*) atuando como classificadores da categoria supervisionada para a previsão de tendência das séries avaliadas. Estas técnicas isoladas (TI) utilizam os 6 atributos (Preço Abertura, Preço Máximo, Preço Mínimo, Preço Fechamento, Quantidade Negociado e Volume Financeiro) já normalizado com Log Retorno e o atributo alvo *ClasseY*.

Já o algorimto RBM, que é da categoria não-supervisioda, recebe somente os mesmos 6 atributos (Preço Abertura, Preço Máximo, Preço Mínimo,Preço Fechamento, Quantidade Negociado e Volume Financeiro) normalizado com Log Retorno, e realiza a redução de dimensionalidade, através de seleção e extração de características latentes e ocultas destes atributos. Como resultado é produzindo uma saída reduzida de 4 novos atributos (características), que serão utilizados como entrada para as 4 TI (*SVM, AD J8, NB* e *KNN*) de categoria supervisionada.

Desta forma, combinando o algoritmo RBM de *deep learning* de categoria não-supervisionada com os 4 TI (*SVM, AD J8, NB* e *KNN*) de categoria supervisionada, propõe-se mais 4 técnicas combinadas (TC) (*RBM + SVM, RBM + AD J8, RBM + NB* e *RBM + KNN*), totalizando 8 propostas de técnicas / modelos de previsão de tendências.

Para cada um dos 8 modelos de previsão de tendências propostos, serão gerados uma nova série temporal binária contendo os sinais ($0 = caiu$ e $1 = subiu$) de saída da predição de tendências dos períodos avaliados, para utilização como indicadores de tendência para a próxima etapa de estratégias de operação.

Durante a avaliação dos modelos de previsão de tendências, serão utilizados dois tipos de janelas de classificação durante as fases de aprendizado (fixa e deslizante) e (treinamento e Validação) nas séries temporais. Estas janelas serão configuradas no Capítulo 5 para verificar qual é a mais adequada em relação aos dados avaliados.

4.4 Estratégias de Operação

As estratégias de operação foram modeladas com a finalidade de realizar uma combinação de ordens de compra e venda (gatilhos) dos ativos avaliados, com operações iniciando e finalizando no mesmo dia (*daytrade*), utilizando uma granularidade de dados diária.

Cada um dos modelos de previsão de tendências ao final do processo de aprendizado, geram um arquivo de dados com as previsões encontradas pelos algoritmos de aprendizado de máquina contendo valores binários (0 e 1). Caso o valor seja igual a 1, significa que o modelo de previsão de tendência está indicando que o preço irá subir, ou seja, a diferença entre o preço de fechamento do dia seguinte (D+1) e o preço de abertura do dia seguinte (D+1) será maior do que 0.

Após receber a série de predição de cada um dos 8 modelos de previsão de tendências propostos, Figura 3, são executadas 8 estratégias de operação que serão detalhadas a seguir.

Algoritmo 1: PSEUDOCÓDIGO DA ESTRATÉGIA PRINCIPAL

Entrada: $SinalPredicao, QtdeDiasPredicao, QtdeClassificadores$
Saída: Arquivo $LogOperacao$ de cada um dos 8 modelos de classificação propostos.
início
    para *i* < $QtdeClassificadores$ **faça**
        zerar ($LogOperacao[i, j]$)
        para *j* < $QtdeDiasPredicao$*[i]* **faça**
            if $SinalPredicao$*[j] = 1* **then**
                comprar ($PrecoAbertura(D + 1)$);
                vender ($PrecoFechamento(D + 1)$);
                calcular ($LogOperacao[i, j]$);
            end
        fim
        gravar ($LogOperacao[i, j]$)
    fim
fim
retorna $LogOperacao$

O Algoritmo 1 **Estratégia principal** executa o gatilho caso o sinal da predição da série temporal avaliada gerada pelo modelo de previsão seja igual a 1, realiza-se a ordem de compra com o valor do Preço de Abertura do dia seguinte (D+1), e a ordem de venda com o valor do Preço de Fechamento do dia seguinte (D+1). Caso o sinal de predição seja igual a 0, o gatilho será cancelado.

Algoritmo 2: PSEUDOCÓDIGO DA ESTRATÉGIA COMPRAR SEMPRE

Entrada: $SinalPredicao, QtdeDiasPredicao, QtdeClassificadores$
Saída: Arquivo $LogOperacao$ de cada um dos 8 modelos de classificação propostos.
início
    para *i* < $QtdeClassificadores$ **faça**
        zerar ($LogOperacao[i, j]$)
        para *j* < $QtdeDiasPredicao$*[i]* **faça**
            comprar ($PrecoAbertura(D + 1)$);
            vender ($PrecoFechamento(D + 1)$);
            calcular ($LogOperacao[i, j]$);
        fim
        gravar ($LogOperacao[i, j]$)
    fim
fim
retorna $LogOperacao$

O Algoritmo 2 **Estratégia Comprar Sempre** executa o gatilho indiferente do sinal da série de predição, será realizado uma ordem de compra com o valor do Preço de Abertura e uma ordem de venda com o valor do Preço de Fechamento todos os dias da série avaliada.

O Algoritmo 3 **Estratégia *Buy and Hold*** será realizado somente uma ordem de compra com o valor do Preço de Abertura do 1º dia (D1) e uma ordem de venda com o valor do Preço de Fechamento do último dia (DN) da série avaliada.

Algoritmo 3: PSEUDOCÓDIGO DA ESTRATÉGIA *Buy and Hold*

Entrada: $SinalPredicao, QtdeDiasPredicao, QtdeClassificadores$
Saída: Arquivo $LogOperacao$ de cada um dos 8 modelos de classificação propostos.
início
    para *i* < $QtdeClassificadores$ **faça**
        comprar ($PrecoAbertura(D1)$);
        vender ($PrecoFechamento(DN)$);
        calcular ($LogOperacao[i, j]$);
        gravar ($LogOperacao[i, j]$)
    fim
fim
retorna $LogOperacao$

O Algoritmo 4 **Estratégia Aleatória** executa o gatilho caso o resultado da predição da série temporal avaliada gerada pelo modelo de previsão seja igual a 1, executado um algoritmo que gera uma semente aleatória binária (0,1) e caso o resultado seja igual a 1, realiza-se a ordem de compra com o valor do Preço de Abertura do dia seguinte (D+1) e, a ordem de venda com o valor do Preço de Fechamento do dia seguinte (D+1). Caso contrário, o gatilho será cancelado.

Algoritmo 4: PSEUDOCÓDIGO DA ESTRATÉGIA ALEATÓRIA

Entrada: $SinalPredicao, QtdeDiasPredicao, QtdeClassificadores$
Saída: Arquivo $LogOperacao$ de cada um dos 8 modelos de classificação propostos.
início
    para *i* < $QtdeClassificadores$ **faça**
        zerar ($LogOperacao[i, j]$)
        para *j* < $QtdeDiasPredicao$*[i]* **faça**
            rodar ($SementeRandomica$) **if** $SementeRandomica$ = *1* **then**
                comprar ($PrecoAbertura(D + 1)$);
                vender ($PrecoFechamento(D + 1)$);
                calcular ($LogOperacao[i, j]$);
            end
        fim
        gravar ($LogOperacao[i, j]$)
    fim
fim
retorna $LogOperacao$

O Algoritmo 5 **Estratégia Alvos (0,5%, 1,0% e 2,0%)** executa o gatilho caso o sinal da predição da série temporal avaliada gerada pelo modelo de previsão seja igual a 1, realiza-se a ordem de compra com o valor do Preço de Abertura do dia seguinte (D+1) e, a ordem de venda com o valor do Preço de Abertura do dia seguinte (D+1) reajustado com cada um

dos 3 (três) tipos de alvos simultâneos. Caso o sinal de predição seja igual a "0", o gatilho será cancelado.

Algoritmo 5: PSEUDOCÓDIGO DAS ESTRATÉGIAS ALVOS 0,50%, 1,00% E 2,00%

Entrada: $SinalPredicao$, $QtdeDiasPredicao$, $QtdeClassificadores$
Saída: Arquivo $LogOperacao$ de cada um dos 8 modelos de classificação propostos.
início
  para *i* < $QtdeClassificadores$ **faça**
    zerar ($LogOperacao[i, j]$)
    para *j* < $QtdeDiasPredicao$*[i]* **faça**
      if $SinalPredicao$*[j] = 1* **then**
        comprar ($PrecoAbertura(D + 1)$);
        vender ($PrecoAbertura(D + 1)$ * $reajuste$);
        calcular ($LogOperacao[i, j]$);
      end
    fim
    gravar ($LogOperacao[i, j]$)
  fim
fim
retorna $LogOperacao$

O Algoritmo 6 **Estratégia Inversa** executa o gatilho caso o sinal da predição da série temporal avaliada gerada pelo modelo de previsão seja igual a 0, realiza-se a ordem de compra com o valor do Preço de Abertura do dia seguinte (D+1) e, a ordem de venda com o valor do Preço de Fechamento do dia seguinte (D+1). Caso o sinal de predição seja igual a "1", o gatilho será cancelado. Ou seja, esta estratégia é inversa à estratégia principal.

Algoritmo 6: PSEUDOCÓDIGO DA ESTRATÉGIA INVERSA

Entrada: $SinalPredicao$, $QtdeDiasPredicao$, $QtdeClassificadores$
Saída: Arquivo $LogOperacao$ de cada um dos 8 modelos de classificação propostos.
início
  para *i* < $QtdeClassificadores$ **faça**
    zerar ($LogOperacao[i, j]$)
    para *j* < $QtdeDiasPredicao$*[i]* **faça**
      if $SinalPredicao$*[j] <> 1* **then**
        comprar ($PrecoAbertura(D + 1)$);
        vender ($PrecoFechamento(D + 1)$);
        calcular ($LogOperacao[i, j]$);
      end
    fim
    gravar ($LogOperacao[i, j]$)
  fim
fim
retorna $LogOperacao$

Ao final da avaliação da estratégia de operação proposta, para cada código de ativo, serão armazenados os dados e parâmetros que fazem parte do processo de execução das ordens de compra e vendas (gatilhos), como por exemplo, a quantidade de gatilhos e o percentual (%) (ganho, perda e retorno) acumulados no arquivo contendo a *Log* de operação, conforme destacado nos Algoritmos 1, 2, 3, 4, 5 e 6.

4.5 Análise de Resultados

A Análise Estatística apresenta a caracterização da amostra de cada de ativo sendo 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 ativo;
- Preço Médio do ativo;
- Preço Final do ativo;
- Quantidade de dias em que ocorreram altas e baixas; e
- Percentual de dias em que ocorreram altas e baixas.

A análise de desempenho realiza a avaliação dos modelos de aprendizado de máquina 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 1);
- Revocação (Equação 3);
- *F1-score* (Equação 4);
- Precisão (Equação 2); e
- Especificidade (Equação 5).

A análise de operação / risco é responsável por apresentar as melhores combinações encontradas, através da utilização das etapas da metodologia proposta, para cada código de ativo avaliado, comparando os seguintes resultados:

- Percentuais (%) acumulados de (Perda, Ganho e Retorno Total);
- *Buy And Hold*: comprando no início da série e vendendo no final do período avaliado;
- Ibovespa: é o índice da Bolsa de Valores de São Paulo, sendo um indicador de desempenho das ações negociadas na Bovespa;
- Volatilidade: indicator mede a diferença entre os preços máximos e mínimos, utilizada para indicar os topos e fundos do mercado cambial;
- Índice *Sharpe*: avalia a rentabilidade e o risco de um investimento, verificando o quanto de retorno excedente em relação a um ativo livre de risco é compensado através de seu nível de risco; e
- Índice *Sortino*: avalia a rentabilidade e o risco de um investimento utilizando volatilidade das perdas em uma série temporal.

O arcabouço implementado avalia a viabilidade, o grau de risco de investimento das propostas de estratégias de investimento utilizando técnicas de *machine learning* e a possibilidade de aplicação futura destas estratégias em um cenário real no mercado financeiro.

4.6 Considerações finais

Este Capítulo detalhou a etapa de caracterização e transformação de dados, mostrando a forma de extração de dados e os tipos de atributos que serão normalizados, juntamente com a criação do indicador alvo (classeY). Em seguida, foram apresentados os modelos classificadores de previsão de tendências, as estratégias de operação e a descrição das 3 formas de análise de resultados.

Todas as etapas (Coletar, Transformar, Classificar, Operar e Analisar) descritos na metodologia foram implementadas na linguagem de programação *Python*, com a utilização de recursos e funções da biblioteca *Scikit-learning* específica para aplicações de algoritmos de aprendizado de máquina, constituindo um arcabouço automatizado para a avaliação de algoritmos de aprendizado de máquina e estratégias de investimentos para o mercado de ações.

Capítulo 5

Experimentos: Resultados e Análises

Neste Capítulo serão executadas todas as 5 etapas da metodologia proposta Figura 3 (Extração de Dados, Transformação de Dados, Modelos de Previsão de Tendências, Estratégias de Operação e Análise de Resultados) já detalhadas no Capítulo 4, através de um arcabouço implementado avaliando diversos cenários e algumas estratégias de investimentos.

Inicialmente, no experimento de treinamento serão avaliados os critérios de seleção e escolha dos ativos, a configuração de parâmetros com a calibração dos algoritmos, a validação da janela de classificação, escolha do modelo de previsão de tendências e a viabilidade dos custos operacionais.

Por último, no experimento de validação serão apresentados, para cada um dos ativos e índices selecionados, a tabela consolidada contendo dados (desempenho, operação / risco e estatístico) e 3 gráficos: série como o Preço de Fechamento, série de evolução de capital acumulado (Retornos líquido e bruto e custo operacional), *boxplot* do retorno financeiro e das medidas de desempenho.

Tabela 6 – Variação acumulada de indicadores de *Baselines* do mercado brasileiro.

Períodos	Data inicial	Data final	Dias úteis	CDI	SELIC	IBOVESPA
Treinamento	04/11/2010	13/11/2014	1.000	37,96%	38,30%	-18,94%
Validação	05/01/2015	24/01/2019	1.000	41,95%	42,02%	76,74%

A Tabela 6 apresenta o *baseline* dos períodos que serão utilizados nas etapas de Treinamento (2010 a 2014) e Validação (2015 a 2019), contendo alguns indicadores de mercado (CDI, SELIC e Ibovespa). Para os dados que serão utilizados na etapa de Trenamento o Ibovespa apresentou um valor acumulado negativo de -18,95% e, para a fase de Validação, apresentou um valor acumulado positivo de 75,74%. Os dados são de granularidade diária, contendo 1.000 dias úteis, 4 anos de dados diários para cada etapa (Treinamento e Validação), sendo um total de quase 9 anos (2010 a 2019) de dados diários coletados para as simulações.

Tabela 7 – Descrição dos 50 ativos da B3 que serão utilizados para as fases de Treinamento (2013 a 2014) e Validação (2015 a 2018) dos modelos de previsão de tendências.

Ativo	Nome	Setor	Referência (www.*)	Preço Inicial	Preço Final	Variação
BRFS3	BRF SA	Alimentação	brf-global.com	R$ 61,80	R$ 24,19	39,14%
JBSS3	JBS	Alimentação	jbs.com.br	R$ 10,51	R$ 14,47	137,68%
MRFG3	MARFRIG	Alimentação	marfrig.com.br	R$ 5,19	R$ 5,87	113,10%
EMBR3	EMBRAER	Aviação	embraer.com.br	R$ 23,94	R$ 19,67	82,16%
GOLL4	GOL	Aviação	gol.com.br	R$ 15,15	R$ 23,43	154,65%
CMIG4	CEMIG	Energia	cemig.com.br	R$ 12,50	R$ 14,49	115,92%
ELET3	ELETROBRAS	Energia	eletrobras.com.br	R$ 5,59	R$ 33,98	607,87%
ELET6	ELETROBRAS	Energia	eletrobras.com.br	R$ 8,11	R$ 36,95	455,61%
ENBR3	ENERGIAS BR	Energia	energias br.com.br	R$ 8,55	R$ 16,54	193,45%
BBAS3	BANCO BRASIL	Financeiro	bb.com.br	R$ 22,94	R$ 48,95	213,38%
BBDC3	BRADESCO	Financeiro	bradesco.com.br	R$ 34,88	R$ 38,68	110,89%
BBDC4	BRADESCO	Financeiro	bradespar.com.br	R$ 35,94	R$ 43,15	120,06%
BRAP4	BRADESPAR	Financeiro	bradespar.com.br	R$ 14,11	R$ 35,00	248,05%
CIEL3	CIELO	Financeiro	cielo.com.br	R$ 37,84	R$ 10,68	28,22%
ITSA4	ITAUSA	Financeiro	itausa.com.br	R$ 9,45	R$ 13,20	139,68%
ITUB4	ITAU UNIBANCO	Financeiro	itau.com.br	R$ 35,00	R$ 37,49	107,11%
MRVE3	MRV	Imobiliário	mrv.com.br	R$ 7,35	R$ 14,85	202,04%
BRML3	BR MALLS PAR	Imobiliário	brmalls.com.br	R$ 15,92	R$ 14,26	89,57%
CYRE3	CYRELA REALT	Imobiliário	cyrela.com.br	R$ 11,08	R$ 17,35	156,59%
CCRO3	CCR SA	Infra-estrutura	ccr.com.br	R$ 14,57	R$ 14,28	98,01%
ECOR3	ECORODOVIAS	Infra-estrutura	ecorodovias.com.br	R$ 9,88	R$ 11,17	113,06%
VALE3	VALE	Mineração	vale.com.br	R$ 22,32	R$ 56,15	251,57%
BRKM5	BRASKEM	Petroquímico	braskem.com.br	R$ 16,05	R$ 49,45	308,10%
PETR3	PETROBRAS	Petroquímico	petrobras.com.br	R$ 8,30	R$ 29,46	354,94%
PETR4	PETROBRAS	Petroquímico	petrobras.com.br	R$ 8,56	R$ 25,54	298,36%
CSAN3	COSAN	Saneamento	cosan.com.br	R$ 25,20	R$ 43,44	172,38%
SBSP3	SABESP	Saneamento	sabesp.com.br	R$ 17,06	R$ 43,18	253,11%
HYPE3	HYPERA	Saude	hypera.com.br	R$ 15,97	R$ 32,47	203,32%
FLRY3	FLEURY	Saúde	fleury.com.br	R$ 15,16	R$ 22,00	145,12%
QUAL3	QUALICORP	Saúde	qualicorp.com.br	R$ 27,49	R$ 15,85	57,66%
RADL3	RAIADROGASIL	Saúde	rd.com.br	R$ 25,32	R$ 62,29	246,01%
UGPA3	ULTRAPAR	Serviços	ultra.com.br	R$ 49,75	R$ 57,71	116,00%
MULT3	MULTIPLAN	Serviços	multiplan.com.br	R$ 47,36	R$ 25,10	53,00%
WEGE3	WEG	Serviços/Comércio	weg.net	R$ 30,58	R$ 18,82	61,54%
IGTA3	IGUATEMI	Serviços/Comércio	iguatemi.com.br	R$ 24,13	R$ 41,81	173,27%
RENT3	LOCALIZA	Serviços/Locação	localizahertz.com	R$ 34,61	R$ 32,22	93,09%
BBSE3	BBSEGURIDADE	Serviços/Seguro	bbseguridaderi.com.br	R$ 30,83	R$ 29,99	97,28%
CVCB3	CVC BRASIL	Serviços/Turismo	cvc.com.br	R$ 14,88	R$ 61,99	416,60%
CSNA3	SID NACIONAL	Siderúrgica	csn.com.br	R$ 5,37	R$ 10,37	193,11%
GGBR4	GERDAU	Siderúrgica	gerdau.com.br	R$ 9,95	R$ 15,99	160,70%
GOAU4	GERDAU MET	Siderúrgica	gerdau met.com.br	R$ 11,45	R$ 7,54	65,85%
USIM5	USIMINAS	Siderúrgica	usiminas.com	R$ 4,85	R$ 9,80	202,06%
TOTS3	TOTVS	Software	totvs.com	R$ 32,81	R$ 33,60	102,41%
TIMP3	TIM PART S/A	Telecomunicações	tim.com.br	R$ 11,42	R$ 12,55	109,89%
VIVT4	TELEFÔNICA	Telecomunicações	telefonica.com.br	R$ 45,98	R$ 47,77	103,89%
BTOW3	B2W DIGITAL	Varejo	ri.b2w.digital	R$ 21,39	R$ 45,61	213,23%
LAME4	LJ. AMERICANAS	Varejo	americanas.com	R$ 16,95	R$ 20,46	120,71%
LREN3	LJ. RENNER	Varejo	lojasrenner.com.br	R$ 74,88	R$ 44,44	59,35%
PCAR4	P.ACUCAR-CBD	Varejo	gpa.com.br	R$ 96,00	R$ 94,78	98,73%
HGTX3	CIA HERING	Varejo/Vestuário	cia hering.com.br	R$ 18,80	R$ 29,45	156,65%

O índice Ibovespa é composto em média pelos 76 ativos que apresentam uma maior representatividade em relação ao volume negociado no mercado de ações e comporem diferentes setores da economia, tais como: alimentação, aviação, energia, financeiro, imobiliário, infra-estrutura, mineração, petroquímico, saneamento, saúde, serviços, siderúrgica, telecomunicações e varejo.

Como critério de seleção, ao consultar o site da B3, verificou-se que de todos os 76 ativos que compunham o Ibovespa em novembro de 2019, somente 50 tiverem ocorrências de negociações em todos os meses no período de 2010 a 2019, os quais foram relacionados na Tabela 7, ordenados por setor de mercado, já calculado o percentual de variação anual (*buy and hold*) das etapas de treinamento (2010 a 2014) e validação (2015 a 2018), evitando, desta forma, que os resultados sejam enviesados.

5.1 Simulação do Treinamento dos Modelos

A etapa de treinamento compreende o período de 07 e janeiro de 2013 a 18 de agosto de 2014, sendo 400 dias úteis, e foram avaliados todos os 50 ativos da Tabela 7 com a finalidade de calibrar, avaliar e selecionar algumas configurações que possam ser mais aderentes às séries temporais e aos tipos de dados, contemplando o seguinte cenário:

- Quantidade de ativos avaliados: 50 ativos da B3 relacionados na Tabela 7;
- Quantidade de papeis negociados por gatilho: 3 lotes (100, 1.000 e 10.000);
- Quantidade de modelos de previsão de tendência: um total de 8 modelos (*SVM, KNN, AD J48, NB, RBM + SVM, RBM + KNN, RBM + AD J48 e RBM + NB*); e
- Período: 04/11/2010 a 13/11/2014 (1.000 dias úteis), conforme Tabela 6.

Para cada simulação realizada de 1.000 dias úteis na etapa de Treinamento avaliou-se 50 ativos, 3 quantidades de papeis negociados, 8 modelos de previsão de tendências, totalizando uma amostra de 1.248 combinações como um cenário de possibilidades de escolhas de configurações, sem contar com os parâmetros específicos dos algoritmos que serão tratados a seguir.

5.1.1 Avaliação da calibração dos algoritmos

Os algoritmos *Naive Bayes (NB)*, Máquina de Suporte de Vetores *(SVM)*, Árvore de Decisão (AD J48), *K-Nearest Neighbour (KNN)* e Máquina Restita de *Boltzmann (RBM)* de aprendizado de máquina possuem alguns parâmetros a serem calibrados que podem influenciar diretamente no resultado da classificação, que serão detalhados a seguir:

- *NB*: não possui parâmetros para calibração;

- *SVM*: avaliou-se duas diferentes funções *kernel* (k) a *Radial Basis Function (RBF)* e a *polynomial (Poly)* com uma quantidade de interação (qi). Na função *RBF* devemos calibrar a largura da curva *gaussiana* (s) e o custo (c), que define a suavização das margens do hiperplano divisor. Já na função *Poly* basta ajustar a ordem do polinômio (op) usado;
- *AD J48*: tem-se o N que indica o número mínimo de elementos permitidos em cada folha e o C, que é o *confidence* e escolhe o fator de confiança inicial para podar a árvore;
- *KNN*: avaliou-se diversos valores para o número de vizinhos próximos (k) e tamanho de folha (TF) com ponderação de pesos uniforme e distância euclidiana entre os pontos; e
- *RBM*: avaliou-se diversos valores variando a quantidade de componentes (c), a quantidade de épocas(q) e a taxa de aprendizado (lr) com 10 camadas ocultas e a inicialização da rede de forma determinística.

Tabela 8 – Intervalo de parâmetros de calibração dos algoritmos.

Algoritmos	**Parâmetros**	**Variação**	**Resultado da Calibragem**
NB	-	-	-
SVM	qi, s, k	[1;7], [100;500], [$Poly, Rbf$]	qi=200, s=0.3, k=Rbf
AD J48	N, C	[1; 5], [0.01; 0.25]	N=4, C=0.20
kNN	k, TF	[0;50], [0;100]	k=5, TF=30
RBM	c, q, lr	[2;10], [100;500], [0.01;0.15]	c=4, q=200, lr=0.10

Neste sentido, após várias simulações com alternância de diversos parâmetros com o objetivo de ajustar os algoritmos aos dados em questão, na fase de treinamento, os mesmos foram calibrados de acordo com a Tabela 8, na qual a coluna (Resultado da Calibragem) apresenta a configuração final utilizada nos próximos experimentos.

5.1.2 Avaliação da janela de classificação

Durante as simulações foram utilizados 2 tipos de janelas de classificação, no qual variou-se a janela deslizante entre 10 e 45 dias, obtendo a configuração de 20 dias como melhor parâmetro encontrado em relação ao percentual de retorno financeiro acumulado no período avaliado.

Já para a janela fixa, utilizou-se 3 variações (70% / 30%, 80% / 20% e 85% / 15%) para os dados (Treinamento/Teste) respectivamente, obtendo a configuração de 70% / 30% como melhor parâmetro encontrado, utilizando como peso, o percentual de precisão do modelo de previsão de tendências acima de 50%.

No entanto, ao comparar o melhor resultado de configuração entre ambas janelas (deslizante e fixa), a janela de classificação que apresentou os melhores percentuais de precisão

acima de 50% para os 8 modelos de previsão de tendências proposto foi a janela deslizante contendo 20 dias (Treinamento) e 1 dia (Teste), contrariando os vários exemplos encontrados nos trabalhos relacionados ((LIU, 2017) (LEE, 2017) (KIMURA, 2017) (KLABJAN, 2016) (XIONG, 2016) (JIA, 2016) (HRASKO, 2015) (LIU, 2015) (LIUY, 2015) (CAO, 2015) (SHARANG, 2015)) a partir da Revisão Sistemática de Literatura no Capítulo 2.

Com a finalidade de filtrar e diminuir o espaço de exploração de possibilidades de combinações, foram descartada a utilização da janela de classificação fixa com 70% (Treinamento) e 30% de (Teste), ou seja, serão utilizou-se somente a janela deslizante de 20 dias (Treinamento) e 1 dia (Teste) para as demais simulações.

5.1.3 Avaliação dos modelos de previsão de tendências

Para escolher o melhor modelo de previsão de tendências para os dados de Treinamento dos 50 ativos, foi realizada a avaliação dos 8 modelos propostos e consolidados seus respectivos resultados em 4 gráficos de *boxplot* em relação à Precisão (%) e Retorno financeiro líquido para cada um dos 8 modelos e também por 2 grupos de técnicas (isolada (TI) e composta (TC)).

A Figura 4 mostra que as 4 técnicas combinadas (TC) (*RBM + SVM, RBM + AD J8, RBM + NB* e *RBM + KNN*), geraram um resultado melhor do que as técnicas isoladas (TI) (*SVM, AD J8, NB* e *KNN*), em relação a Precisão, Acurácia e ao Retorno financeiro líquido. O modelo de previsão de tendência proposto *RBM + NB* encontrou os melhores resultados em relação às medidas de Acurácia (49,00% a 54,00%) e Retorno Financeiro Líquido (30,00% a 80,00%). Já em relação à medida de Precisão (51,00% a 55,00%) os melhores resultados foram do modelo *RBM + KNN*.

Como critério de seleção, para os 8 modelos de previsão de tendências propostos, consolidou-se todos os Retornos Financeiros Brutos em (%) dos 50 ativos avaliados para o cálculo das médias destes retornos. Conforme destacado na Tabela 9, o modelo de previsão de tendência *RBM + NB* gerou o melhor retorno financeiro Bruto médio no valor de 75,09% e, apresentando um ganho de praticamente 45% superior em relação aos demais modelos de previsão de tendências propostos.

Em tratando de séries temporais de dados de mercado de ações, não adianta utilizar um modelo de previsão de tendências que tenha uma excelente medida de Precisão, se o que é mais importante é o Retorno Financeiro. Desta forma, para as demais simulações somente será utilizado o modelo de previsão de tendência *RBM + NB*, pois obteve o melhor valor de Retorno Financeiro médio entre os 50 ativos para esta série avaliada de 1.000 dias úteis.

Tabela 9 – Fase de Treinamento: 50 ativos - Retorno Financeiro Bruto (%) - Critério de seleção entre os 8 modelos de previsão de tendências propostos.

Ativos	SVM	NB	KNN	AD J48	RBM + SVM	RBM + NB	RBM + KNN	RBM + AD J48
BRFS3	62,44	75,99	81,63	57,70	47,17	76,16	78,91	112,59
JBSS3	88,83	76,63	92,40	74,52	43,87	92,34	34,26	80,16
MRFG3	-11,40	-5,33	-6,77	-47,56	2,43	99,40	69,18	28,60
EMBR3	66,79	65,39	78,03	81,50	73,50	131,71	75,29	44,74
GOLL4	9,27	-95,43	-52,96	-127,32	-57,56	52,90	-38,65	-54,37
CMIG4	39,68	22,44	-6,66	-13,36	49,10	88,56	40,22	6,74
ELET3	-5,96	-27,26	-15,22	-75,86	9,23	2,36	-2,14	-63,73
ELET6	-9,82	15,33	-62,22	-41,26	15,72	49,18	-20,72	-44,71
ENBR3	127,70	154,53	141,01	82,18	55,66	122,99	130,74	122,39
BBAS3	71,14	37,97	78,98	25,16	38,76	41,17	54,85	12,83
BBDC3	102,16	89,68	105,66	80,73	96,43	126,15	99,15	62,68
BBDC4	33,46	25,96	18,15	30,09	43,08	29,27	48,41	62,85
BRAP4	9,94	-9,74	7,65	-29,80	22,25	42,30	47,20	21,67
CIEL3	86,67	125,54	103,34	130,80	79,25	89,09	111,59	74,09
ITSA4	47,77	37,40	35,14	27,67	51,31	38,94	54,58	54,15
ITUB4	24,31	-10,17	-29,10	25,20	3,67	13,58	-17,69	26,18
MRVE3	95,84	41,24	39,20	-25,27	37,02	85,71	74,40	37,05
BRML3	33,18	66,06	-24,90	-4,09	79,60	129,37	27,30	6,94
CYRE3	38,74	-7,13	12,68	-62,38	7,91	55,52	13,16	2,79
CCRO3	39,80	42,40	17,38	75,12	47,76	61,46	71,74	78,03
ECOR3	46,52	30,33	6,26	23,69	41,29	54,65	64,34	40,22
VALE3	-5,96	-42,03	-27,84	-37,98	-37,09	-14,25	-20,05	-39,30
BRKM5	33,86	41,25	56,24	-3,49	33,85	56,25	-14,00	14,61
PETR3	-45,15	-59,40	-21,43	-63,34	46,09	74,47	26,47	-82,70
PETR4	-43,03	-29,94	-23,76	-58,19	-26,09	-1,22	-13,88	-14,68
CSAN3	117,02	95,54	78,41	64,62	86,32	124,09	119,09	79,00
SBSP3	85,42	114,94	85,94	102,54	75,74	124,24	77,28	26,30
HYPE3	44,36	48,77	71,69	-0,88	59,46	100,87	62,45	22,58
FLRY3	-2,97	-6,78	2,47	-24,90	-15,15	-13,50	-14,65	-54,08
QUAL3	68,44	55,33	6,52	-13,86	35,19	86,93	31,23	81,17
RADL3	87,96	74,64	52,36	24,86	92,11	97,36	74,02	36,64
UGPA3	132,92	98,41	91,83	62,87	124,60	156,54	126,68	124,09
MULT3	70,53	51,83	59,45	20,32	72,62	111,80	67,77	56,91
WEGE3	28,32	31,22	40,17	28,10	36,70	36,54	29,05	3,51
IGTA3	37,59	37,38	-5,49	2,72	95,32	66,99	30,38	84,16
RENT3	87,48	79,41	55,48	-12,00	67,62	142,85	44,59	44,68
BBSE3	207,50	221,71	158,03	149,18	196,41	179,40	152,28	172,93
CVCB3	58,96	24,57	67,82	48,90	30,61	86,91	55,30	5,96
CSNA3	-112,88	-48,60	-78,27	-99,61	-35,09	-28,41	2,83	-115,47
GGBR4	-15,32	-11,15	-24,43	4,06	-0,79	1,18	-19,13	-44,60
GOAU4	21,16	49,26	-30,54	34,79	-16,65	27,32	19,65	25,82
USIM5	-42,75	-96,22	-39,35	-155,57	-32,70	-53,26	-32,18	-161,54
TOTS3	48,70	102,14	9,63	-6,22	75,66	89,42	49,10	-6,14
TIMP3	93,21	9,50	46,15	33,70	94,12	104,79	42,56	10,87
VIVT4	39,98	59,02	80,28	-6,39	68,92	87,08	74,23	12,15
BTOW3	68,17	68,23	36,97	45,18	68,04	167,46	93,48	67,83
LAME4	107,88	98,84	49,79	43,41	94,98	138,12	84,76	108,11
LREN3	79,42	103,83	65,61	37,66	83,47	141,60	106,41	38,85
PCAR4	40,64	80,38	51,86	83,38	76,89	119,76	97,89	83,29
HGTX3	5,73	19,13	-21,06	-55,07	22,91	60,59	9,92	6,99
Média	**45,89**	**40,46**	**30,28**	**10,73**	**45,23**	**75,09**	**47,59**	**26,00**

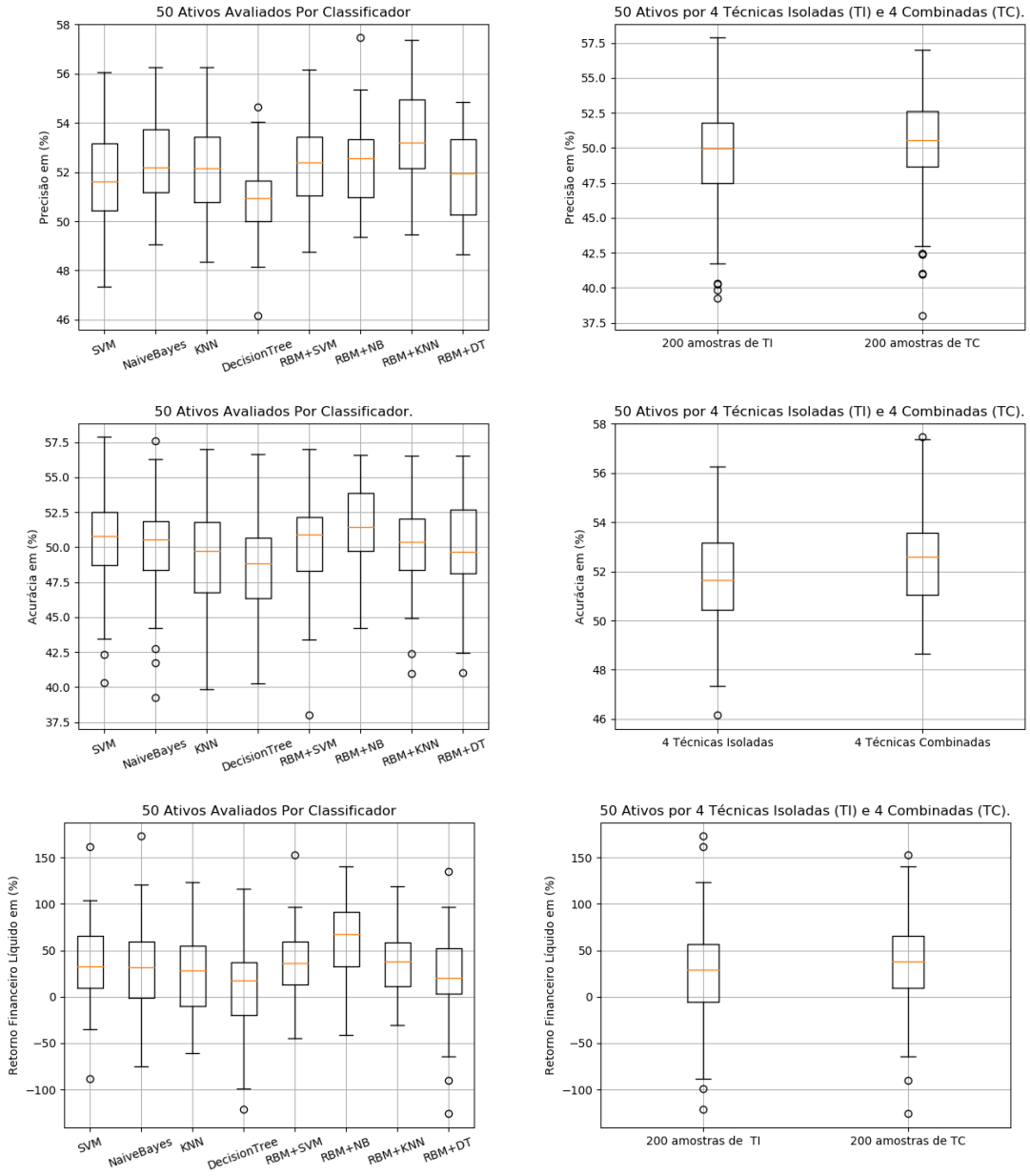


Figura 4 – Treinamento: Avaliação dos 8 modelos propostos para os 50 ativos, período de 1.000 dias úteis de 04/11/2010 a 13/11/2014.

5.1.4 Avaliação da viabilidade dos custos operacionais

A Tabela 10 apresenta os valores dos custos operacionais de cada taxa utilizados durante os experimentos acrescidos de 3 exemplos de lucro bruto (R$100,00, R$1.000,00 e R$10.000,00), no qual fica evidente que quanto maior for o valor investido, menor será o custo operacional, 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 (*daytrade*) utilizada nas estratégias.

Para as simulações realizadas foram contabilizados os custos das operações, contemplando

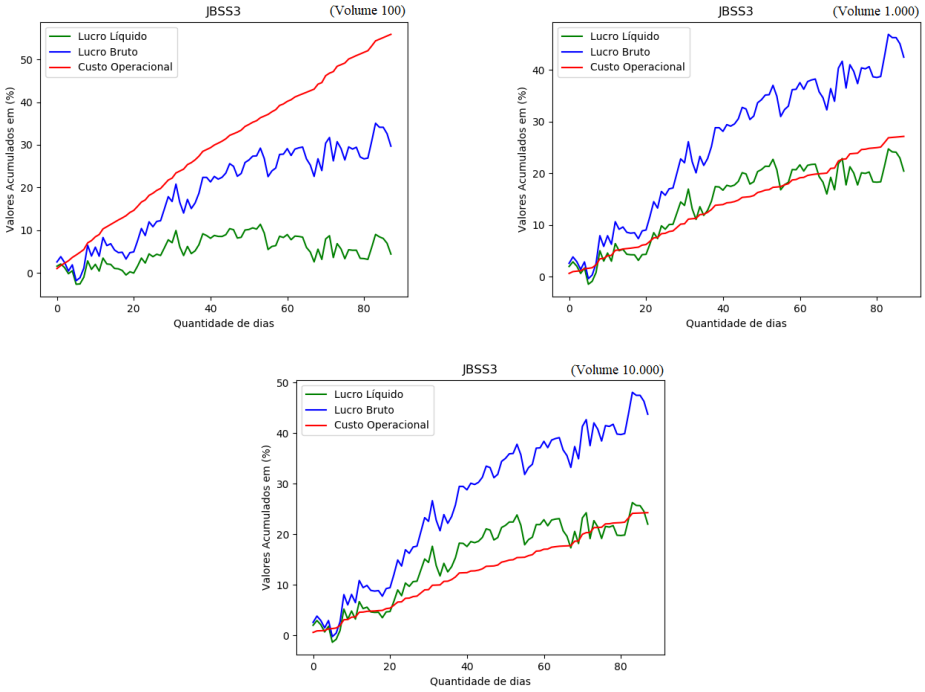


Figura 5 – Comparação entre o custo operacional, o retorno bruto e o retorno líquido acumulado de 132 dias úteis de 01/12/2014 a 16/07/2015.

Tabela 10 – Valores utilizados e exemplos dos 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
Emolumentos (0,025%)	R$ 0,0250	R$ 0,2500	R$ 2,5000
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**

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 a muitas corretoras isentarem seus clientes da taxa de custódia, esta não foi acrescentada nos custos das transações.

A Figura 5, apresenta um comparativo entre a mesma série temporal de 132 dias úteis no período de 01/12/2014 a 16/07/2015 do código de ativo JBSS3, porém contendo lotes de negociação de 100, 1.000 e 10.000 respectivamente, por ordem executada. Fica evidente que o investidor com um capital menor possui um custo operacional maior, o que muitas

vezes, pode inviabilizar uma estratégia de negociação lucrativa. Diante deste fator, foi escolhido para as demais simulações, um volume negociado de 1.000 lotes por estratégia de operação que for executada.

5.2 Simulação da Validação dos Modelos

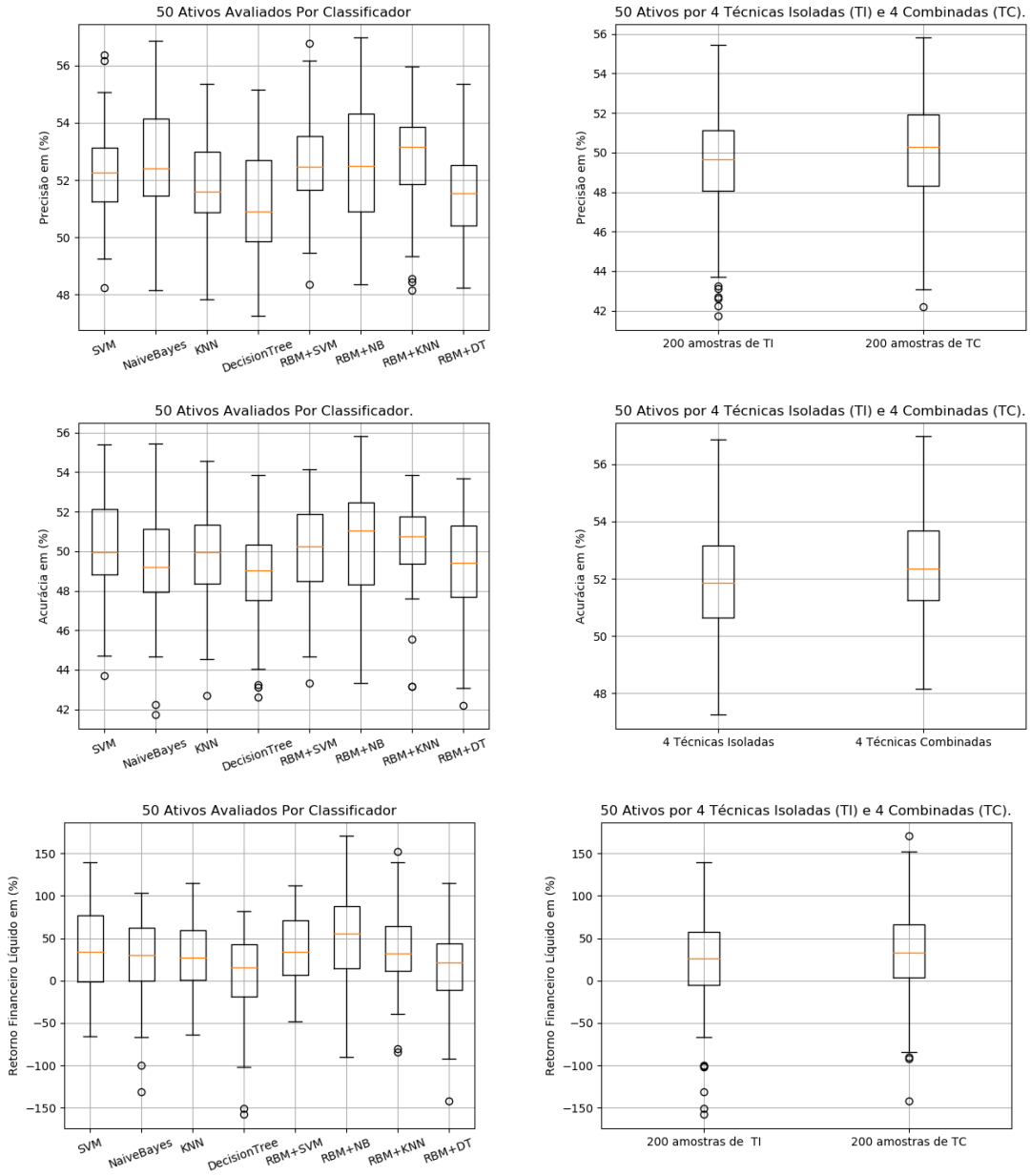


Figura 6 – Validação: Avaliação dos 8 modelos propostos para os 50 ativos, período de 1.000 dias úteis de 05/01/2015 18/11/2018.

A etapa de validação compreende o período de 07 de janeiro de 2015 a 24 de janeiro de 2019, sendo 1.000 dias úteis, inicialmente serão avaliados os 50 ativos da B3 da Tabela 7 e, em seguida, mais 10 índices globais da Tabela 13. Serão apresentados o detalhamento

dos resultados das contendo as tabelas e gráficos produzidos automaticamente durante a simulação pelo arcabouço proposto. Como configurações serão negociados 1 lote 1.000 papéis e avaliados 8 modelos de previsão de tendência (AD J48, NB, RBM + SVM, RBM + KNN, RBM + AD J48 e RBM + NB) já descritos no Capítulo 4.

Tabela 11 – Fase de Validação: 50 ativos - Retorno Financeiro Bruto (%) - Critério de seleção entre os 8 modelos de previsão de tendências propostos.

Ativos	SVM	NB	KNN	AD J48	RBM + SVM	RBM + NB	RBM + KNN	RBM + AD J48
BRFS3	26,18	-38,41	-22,40	-19,28	38,29	70,10	32,95	-16,01
JBSS3	98,57	128,79	142,89	102,33	135,61	166,17	105,52	148,00
MRFG3	96,57	75,71	53,88	-4,17	28,01	104,94	103,61	62,12
EMBR3	-31,70	44,37	-77,20	-66,59	6,99	28,20	4,92	-46,90
GOLL4	-70,72	-168,25	-69,87	-60,90	-26,36	-17,99	-108,42	-85,93
CMIG4	-10,64	-81,77	22,78	-15,85	-36,10	-28,52	5,65	-9,37
ELET3	179,27	57,23	91,17	-36,13	115,21	140,73	178,66	77,71
ELET6	144,38	91,91	148,05	73,22	123,44	192,52	195,26	45,87
ENBR3	140,60	132,49	116,26	42,42	120,47	161,20	87,46	47,37
BBAS3	38,18	39,02	56,70	-27,14	30,19	40,19	29,07	-8,05
BBDC3	106,69	39,13	48,82	32,80	48,74	80,63	72,64	43,91
BBDC4	53,59	44,87	39,41	66,88	52,44	41,85	23,02	57,52
BRAP4	-66,70	-39,14	77,70	21,71	-7,41	-4,56	16,91	28,48
CIEL3	-26,06	5,99	0,19	-23,89	-13,21	10,24	4,65	-14,78
ITSA4	-6,56	-7,17	5,06	-40,04	-27,96	-8,83	41,98	-15,21
ITUB4	6,02	17,41	-8,16	16,20	21,00	27,51	41,40	7,43
MRVE3	131,55	92,77	78,23	68,37	97,97	126,06	111,74	116,63
BRML3	26,64	19,58	18,68	5,78	69,19	82,09	30,23	25,90
CYRE3	63,38	84,26	65,29	28,39	34,78	80,74	53,24	37,78
CCRO3	3,84	34,77	-18,50	-3,48	36,66	82,45	34,55	4,95
ECOR3	-31,70	-44,20	1,55	-42,14	50,52	15,01	28,45	-13,44
VALE3	-36,05	-48,31	-36,78	-52,64	-59,11	-25,20	1,78	-9,07
BRKM5	127,39	126,48	124,43	94,76	141,59	218,45	162,53	105,11
PETR3	67,31	37,93	47,80	-38,03	86,58	67,50	84,82	15,93
PETR4	-39,85	-13,58	-5,51	-130,96	-3,46	-4,41	13,85	-20,40
CSAN3	121,46	104,63	107,60	25,94	110,37	147,97	112,55	54,15
SBSP3	86,06	123,80	118,44	55,36	60,09	116,29	72,40	35,19
HYPE3	66,74	103,38	73,69	99,52	99,10	106,95	72,05	118,05
FLRY3	149,21	80,72	101,20	55,59	143,60	136,91	109,97	96,38
QUAL3	50,00	19,55	-33,03	18,31	38,74	71,05	28,72	-17,44
RADL3	106,23	55,83	61,32	73,06	98,59	61,86	71,55	52,73
UGPA3	17,72	69,60	11,85	16,26	50,73	64,37	25,28	-2,54
MULT3	94,01	58,82	87,02	53,16	78,95	99,89	69,68	79,64
WEGE3	10,86	24,34	11,81	-11,66	22,28	45,20	13,04	-12,12
IGTA3	46,88	49,37	57,66	46,58	52,02	103,00	40,26	63,41
RENT3	100,11	56,97	103,82	104,83	92,77	122,09	121,11	114,92
BBSE3	60,34	95,62	18,68	60,98	10,65	52,16	84,30	60,21
CVCB3	42,84	58,13	43,79	34,82	41,56	104,76	78,03	-18,24
CSNA3	26,62	-78,10	-44,57	-24,22	-11,81	-22,95	-14,17	-54,49
GGBR4	-30,30	-128,03	-81,89	-193,50	-62,28	-115,91	-103,27	-88,43
GOAU4	22,43	-81,02	26,83	-129,50	28,06	-14,53	1,42	-117,42
USIM5	-84,15	-85,41	-45,70	-202,47	-9,48	16,10	-50,97	-181,97
TOTS3	-6,26	-2,79	21,79	30,66	-41,05	50,99	36,19	43,29
TIMP3	43,40	18,32	35,56	55,19	55,24	113,66	49,65	48,92
VIVT4	6,60	14,07	9,68	-4,57	44,39	42,02	27,06	17,54
BTOW3	120,28	124,39	43,96	22,05	35,04	171,83	42,10	29,54
LAME4	43,43	37,75	15,01	26,00	59,68	75,00	8,31	17,81
LREN3	91,39	115,25	85,70	27,77	100,08	140,55	105,86	146,52
PCAR4	-3,60	23,26	-44,88	6,90	-22,94	2,25	-4,16	-51,73
HGTX3	113,96	67,27	35,24	78,65	112,99	76,32	62,28	30,70
Média	**45,73**	**31,15**	**34,42**	**6,35**	**45,03**	**68,30**	**48,31**	**21,00**

A Figura 6 mostra que as 4 técnicas combinadas (TC) (*RBM + SVM, RBM + AD J8, RBM + NB* e *RBM + KNN*), geraram novamente um resultado melhor do que as técnicas isoladas (TI) (*SVM, AD J8, NB* e *KNN*), em relação a Precisão, Acurácia e ao Retorno financeiro

Tabela 12 – Modelo RBM + NB: Treinamento (1.000 dias de 04/11/2010 a 13/11/2014) e Validação (1.000 dias de 07/01/2015 a 24/01/2019).

	Simulação de Treinamento (%)						Simulação de Validação (%)					
Ativos	Distribuição Baixa	Distribuição Alta	Preci.	Buy and Hold	Retorno Bruto	Qtde. Gatilhos	Distribuição Baixa	Distribuição Alta	Preci.	Buy and Hold	Retorno Bruto	Qtde. Gatilhos
BRFS3	48,00	52,00	52,59	127,36	**76,16**	502	55,10	44,90	50,52	-50,51	**70,10**	477
JBSS3	53,00	47,00	52,32	56,78	**92,34**	495	53,00	47,00	51,50	31,27	**166,17**	501
MRFG3	53,80	46,20	51,42	-50,99	**99,40**	492	56,90	43,10	47,97	10,87	**104,94**	469
EMBR3	51,50	48,50	51,92	75,09	**131,71**	495	54,40	45,60	48,37	-14,80	**28,20**	490
GOLL4	59,50	40,50	44,89	-47,12	**52,90**	421	58,00	42,00	44,19	45,36	**-17,99**	473
CMIG4	49,30	50,70	53,29	-46,75	**88,56**	516	54,90	45,10	48,12	13,21	**-28,52**	453
ELET3	58,10	41,90	46,09	-62,46	**2,36**	486	52,90	47,10	50,62	421,53	**140,73**	482
ELET6	54,40	45,60	48,90	-57,60	**49,18**	544	52,90	47,10	50,91	295,16	**192,52**	493
ENBR3	47,80	52,20	55,75	-62,53	**122,99**	565	51,20	48,80	54,56	77,56	**161,20**	515
BBAS3	49,30	50,70	50,53	-22,63	**41,17**	570	52,00	48,00	52,33	94,11	**40,19**	493
BBDC3	51,00	49,00	52,47	16,71	**126,15**	526	48,30	51,70	53,73	9,04	**80,63**	536
BBDC4	49,80	50,20	50,38	-2,40	**29,27**	526	49,50	50,50	49,81	16,65	**41,85**	522
BRAP4	50,60	49,40	50,85	-57,60	**42,30**	531	51,80	48,20	47,09	122,88	**-4,56**	567
CIEL3	47,40	52,60	54,46	138,07	**89,09**	527	54,20	45,80	47,84	-59,57	**10,24**	464
ITSA4	49,50	50,50	51,84	-25,58	**38,94**	488	53,70	46,30	44,56	32,94	**-8,83**	478
ITUB4	50,60	49,40	48,74	-14,39	**13,58**	478	49,90	50,10	50,28	5,90	**27,51**	529
MRVE3	52,40	47,60	53,11	-46,54	**85,71**	450	51,80	48,20	53,24	84,69	**126,06**	494
BRML3	51,70	48,30	54,30	-3,42	**129,37**	477	52,70	47,30	51,95	-8,65	**82,09**	487
CYRE3	52,70	47,30	51,10	-40,08	**55,52**	501	52,30	47,70	51,17	46,97	**80,74**	469
CCRO3	48,90	51,10	53,11	-53,18	**61,46**	531	52,70	47,30	52,43	-1,65	**82,45**	515
ECOR3	52,10	47,90	51,18	-21,25	**54,65**	510	53,30	46,70	46,64	10,84	**15,01**	476
VALE3	52,50	47,50	49,27	-48,98	**-14,25**	481	50,90	49,10	50,57	125,80	**-25,20**	530
BRKM5	51,20	48,80	51,17	7,32	**56,25**	469	53,10	46,90	50,00	172,72	**218,45**	472
PETR3	54,90	45,10	50,20	-46,31	**74,47**	502	53,40	46,60	52,42	211,60	**67,50**	496
PETR4	54,20	45,80	46,00	-41,73	**-1,22**	500	53,20	46,80	48,73	164,64	**-4,41**	513
CSAN3	49,30	50,70	55,07	4,02	**124,09**	503	47,90	52,10	53,85	60,08	**147,97**	507
SBSP3	48,10	51,90	54,33	-43,85	**124,24**	543	49,10	50,90	52,29	127,08	**116,29**	503
HYPE3	54,30	45,70	51,29	-35,34	**100,87**	466	49,50	50,50	52,74	85,75	**106,95**	529
FLRY3	54,80	45,20	46,65	-27,44	**-13,50**	478	52,60	47,40	51,75	37,45	**136,91**	514
QUAL3	53,20	46,80	51,78	24,52	**86,93**	506	51,20	48,80	51,16	-35,14	**71,05**	518
RADL3	51,30	48,70	51,50	182,17	**97,36**	499	51,60	48,40	52,49	121,19	**61,86**	522
UGPA3	48,50	51,50	55,16	99,22	**156,54**	533	50,00	50,00	53,46	13,28	**64,37**	520
MULT3	50,00	50,00	54,55	18,19	**111,80**	550	49,30	50,70	53,44	-39,01	**99,89**	494
WEGE3	52,90	47,10	50,21	27,22	**36,54**	476	53,10	46,90	48,32	-31,92	**45,20**	507
IGTA3	51,00	49,00	51,13	-34,15	**66,99**	530	49,10	50,90	52,82	60,81	**103,00**	515
RENT3	50,50	49,50	55,58	11,28	**142,85**	520	48,80	51,20	52,78	-5,73	**122,09**	540
BBSE3	47,60	52,40	55,89	53,47	**179,40**	526	52,40	47,60	47,76	-2,26	**52,16**	469
CVCB3	50,60	49,40	52,95	199,07	**86,91**	508	48,60	51,40	55,80	262,78	**104,76**	500
CSNA3	57,00	43,00	45,51	-62,43	**-28,41**	457	56,80	43,20	46,68	77,28	**-22,95**	497
GGBR4	56,70	43,30	45,09	-43,23	**1,18**	428	54,90	45,10	43,34	50,38	**-115,91**	503
GOAU4	53,20	46,80	48,88	-42,74	**27,32**	491	56,30	43,70	46,51	-28,34	**-14,53**	473
USIM5	58,00	42,00	44,20	-62,87	**-53,26**	457	58,40	41,60	44,39	84,71	**16,10**	437
TOTS3	52,60	47,40	54,93	-65,11	**89,42**	477	53,00	47,00	48,90	2,00	**50,99**	499
TIMP3	50,90	49,10	49,81	-9,87	**104,79**	528	51,20	48,80	52,47	8,21	**113,66**	507
VIVT4	48,60	51,40	56,60	-14,22	**87,08**	530	50,50	49,50	53,09	3,23	**42,02**	518
BTOW3	54,00	46,00	49,11	-6,10	**167,46**	505	51,40	48,60	52,16	93,98	**171,83**	556
LAME4	51,70	48,30	52,22	-14,90	**138,12**	540	51,10	48,90	52,26	17,19	**75,00**	532
LREN3	51,30	48,70	54,07	6,81	**141,60**	492	48,90	51,10	54,71	-33,74	**140,55**	510
PCAR4	50,20	49,80	54,70	25,33	**119,76**	532	51,30	48,70	50,00	-1,05	**2,25**	528
HGTX3	52,50	47,50	49,67	-10,98	**60,59**	459	53,50	46,50	50,10	47,02	**76,32**	515
Média	**51,86**	**48,14**	**51,34**	**-3,04**	**75,09**	**502,34**	**52,25**	**47,75**	**50,46**	**56,68**	**68,30**	**502,14**

líquido. O modelo de previsão de tendência proposto *RBM + NB* encontrou os melhores resultados em relação às medidas de Acurácia (49,00% a 51,00%), Precisão (51,00% a 54,00%) e também o Retorno Financeiro Líquido (20,00% a 80,00%).

Na Tabela 11 são apresentados os Retornos Financeiros Brutos médios dos 8 modelos de previsão de tendências propostos, com destaque para o modelo *RBM + NB* novamente, que gerou o melhor retorno financeiro Bruto médio no valor de 68,30% e, apresentando um ganho de praticamente 22,00% superior em relação aos demais modelos de previsão de tendências propostos. Este resultado mostra que o modelo *RBM + NB* conseguiu generalizar após a calibração dos diversos parâmetros na etapa de Treinamento, diante de novos dados agora na etapa de Validação. O modelo *RBM + NB* apresentou os melhores resultados em relação aos Retornos Financeiros Brutos médios para os 50 ativos avaliados, tanto nos dados de 1.000 dias úteis da fase de Treinamento, quanto em outro período de 1.000 dias úteis na fase de Validação. Desta forma, os demais resultados, análises e detalhamentos serão somente em relação ao modelo *RBM + NB*.

Na Tabela 12 é apresentado um comparativo entre os resultados obtidos nas fases de Treinamento e Validação do modelo *RBM + NB*, sendo composta dos 50 ativos, distribuição (alta e baixa), Precisão (Preci.), variação no período (*Buy and Hold*, Retorno Financeiro Bruto e Quantidade (Qtde.) de Gatilhos executados. Com destaque na coloração verde estão todos os Retornos Financeiros Brutos que obtiveram valores acima do *baseline* comparativo *Buy and Hold*, ou seja, comprar no primeiro dia e vender somente no último dia da série / período avaliado. Para um total de 50 ativos avaliados, somente 5 na fase Treinamento e 13 na fase de Validação obtiveram resultados inferior em relação ao *Buy and Hold*, ou seja, o modelo *RBM + NB* somando os resultados das duas fases (Trenamento e Validação) tem uma média de ganho de 72,00% das vezes em relação à estratégia de *Buy and Hold*.

Em relação ao resultado da Precisão média do modelo *RBM + NB* foi baixa, nada discrepante, apresentou 51,34% no Treinamento e 50,46 na Validação. Ao analisar os resultados por setor, o detaque foi para o de Alimentação (BRFS3, JBSS3 e MRFG3), que apresentou uma Precião média abaixo de 50,00%, porém obteve o melhor Retorno Financeiro médio de 113,76% entre os setores na etapa de Validação.

No Apêndice A serão apresentados mais detalhes sobre os resultados (Gráficos e Tabelas) produzidos na etapa de Validação dos 50 ativos da B3 avaliados de forma automatizada pelo arcabouço proposto, em relação às distribuições das séries diárias do Preços de Fechamento por ativo, do Retorno Financeiro Acumulado por ativo, do Retorno Financeiro Médio das Estratégias avaliadas e o extrato consolidado dos demais indicadores.

5.2.1 Avaliação de 10 índices globais

Para diversificar e avaliar a robustez do arcabouço proposto e dos seus respectivos modelos de classificação de previsão de tendências, em especial o modelo *RBM + NB*, será realizado outra etapa de Validação (2015 a 2019) contendo dados de novos mercados. São apresentados os 10 índices globais selecionados das principais bolsas de valores mundiais (Austrália, França, Estados Unidos da América, Alemanha, China, Canadá, Japão e Inglaterra), sendo compostos pelas principais empresas representativas mundiais na Tabela 13.

Tabela 13 – Validação: Descrição dos 10 índices globais por setor, *buy and hold*, período de Validação (2015 a 2018) e a distribuição de classes de (Baixa e Alta).

Índice	Nome	Setor / Mercado	Referencias (www.*)	*Buy and hold* 2015 a 2019	Distribuição	
					Baixa	Alta
ASX11	ASX 200	Sydney- Austrália	marketindex.com.au	-9,24%	46,80%	53,20%
CAC11	CAC 40	Euronext Paris - França	live.euronext.com	11,82%	48,40%	51,60%
CCMP1	Nasdaq	New York - Estados Unidos	nasdaq.com	52,51%	43,80%	56,20%
DAX11	DAX 30	Frankfurt - Alemanha	boerse-frankfurt.de	7,72%	48,30%	51,70%
HSI11	HSI	Hong Kong China	hkex.com.hk	5,19%	52,00%	48,00%
NYA11	NYSE	New York - Estados Unidos	nasdaq.com	11,57%	47,10%	52,90%
SHCOM	SSE	Shanghai China	english.sse.com.cn	-27,00%	40,90%	59,10%
SPTSH	TSE 300	Toronto - Canadá	tsx.com	-9,34%	49,70%	50,30%
TPX11	TOPIX	Tokyo - Japão	jpx.co.jp	22,15%	52,80%	47,20%
UKX11	FTSE 100	London	londonstockexchange.com	-10,88%	47,30%	52,70%

Na Tabela 14 são apresentados os Retornos Financeiros Brutos médios dos 8 modelos de previsão de tendências propostos, com destaque para o modelo RBM + NB novamente, que gerou o melhor retorno financeiro Bruto médio no valor de 32,57% e, apresentando um ganho de quase 10,00% superior em relação aos demais modelos de previsão de tendências propostos. Este resultado mostra que o modelo RBM + NB conseguiu generalizar após a calibração dos diversos parâmetros na etapa de Treinamento, diante de novos dados e de mercados mundiais distintos agora na etapa de Validação.

Tabela 14 – Validação: 10 índices globais - Retorno Financeiro Bruto (%) - Critério de seleção entre os 8 modelos de previsão de tendências propostos.

Índices	SVM	NB	KNN	AD J48	RBM + SVM	RBM + NB	RBM + KNN	RBM + AD J48
CCMP1	17,22	25,24	18,25	6,62	24,41	19,71	39,05	16,02
ASX11	59,23	41,78	37,9	25,03	40,79	48,09	36,18	24,67
CAC11	8,34	10,27	6,71	-5,31	-1,95	31,51	2,5	10,16
DAX11	9,89	18,02	8,1	17,16	21,33	21,57	19,99	-2
HSI11	-17,19	-26,51	-27,28	-20,31	-22,04	-9,31	-19,21	-16,07
NYA11	29,13	38,87	16,35	22,72	30,09	32,27	22,74	22,95
SHCOM	136,37	116,68	113,46	106,02	123,63	122,88	130,37	116,45
SPTSH	8,83	-2,39	0,82	-4,25	-6,6	10,94	-2,32	-3,26
TPX11	6,86	13,19	-0,93	-4,26	8,51	-11,55	-16,8	-7,63
UKX11	43,3	26,84	52,44	20,08	40,02	59,57	20,97	37,96
Média	**30,20**	**26,20**	**22,58**	**16,35**	**25,82**	**32,57**	**23,35**	**19,93**

A Figura 7 mostra para este período utilizado na simulação da validação dos modelos, para os 10 índices globais, que as 4 técnicas combinadas (TC) (*RBM + SVM, RBM + AD J8, RBM + NB* e *RBM + KNN*), geraram um resultado melhor do que as técnicas isoladas (TI) (*SVM, AD J8, NB* e *KNN*), em relação a Precisão, Acurácia e ao Retorno financeiro líquido.

Diferente do resultado para os 50 ativos da B3, para estes 10 índices, o modelo de previsão de tendência proposto *RBM + NB* novamente encontrou os melhores resultados em relação ao retorno financeiro líquido variando de 10% a 36% de lucro, destacando-se dos demais modelo de previsão de tendências.

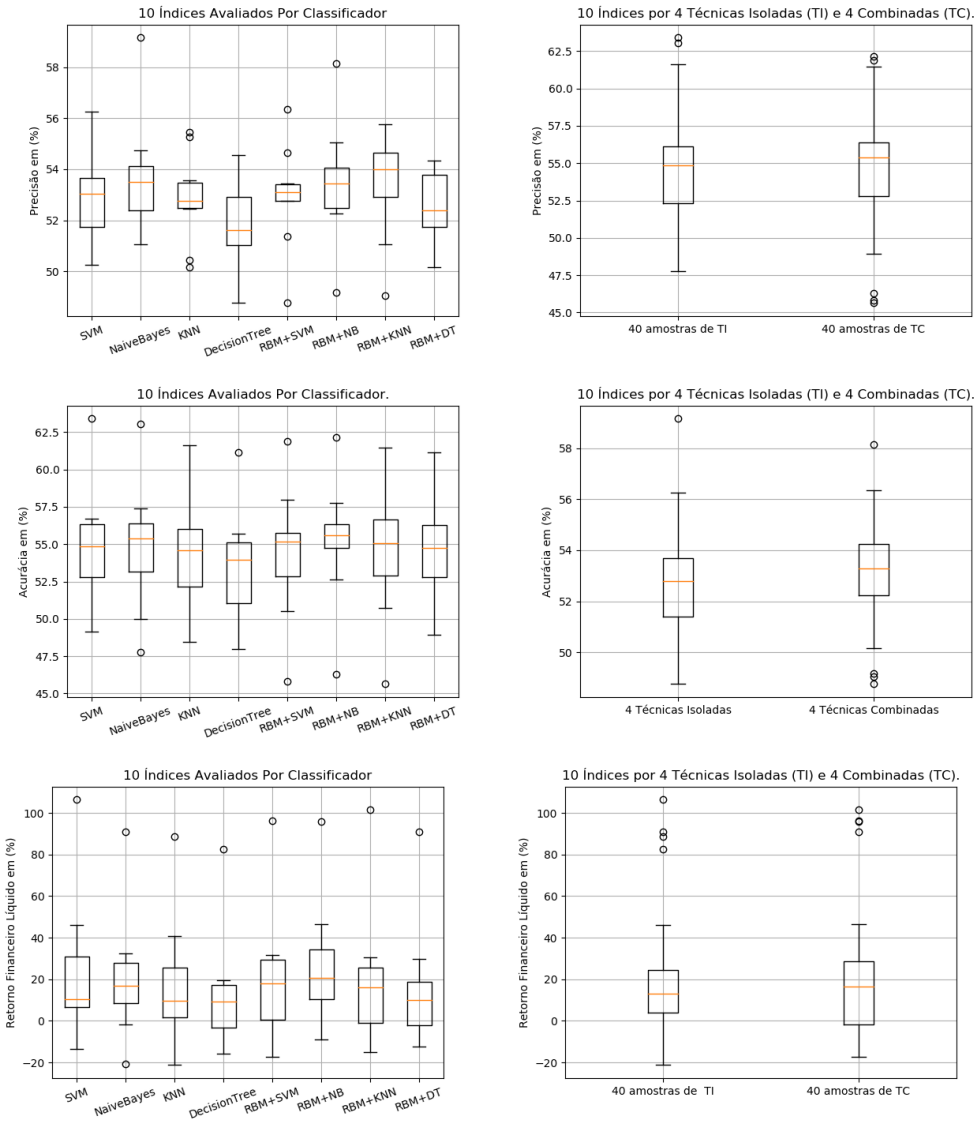


Figura 7 – Validação dos 10 índices globais: comparação entre as medidas de desempenho e o retorno financeiro de 1.000 dias úteis de 05/01/2015 18/11/2018.

Na Tabela 15 são apresentados os resultados detalhados de cada indicador gerado automaticamente durante a simulação de validação dos 10 índices globais do modelo de previsão de tendências selecionado *RBM + NB*, contendo os retornos financeiros em percentual das 7 estratégias de operação (Comprar sempre, Inversa, Aleatória, e os Alvos (0,50%, 1,00% e 2,00%)), conforme já detalhados nas descrições dos Algoritmos 1, 2, 3, 4, 5 e 6.

Os indicadores estatísticos (média, desvio padrão, desvio absoluto, mediana, quantil, má-

ximo, mínimo, amplitude e variância), foram calculados em relação ao retorno financeiro bruto diário de todos os dias em que houve operação pelo modelo proposto, utilizando como referência a estratégia principal do Algoritmo 1.

Tabela 15 – Extrato consolidado da Validação do modelo *RBM + KNN* para os 10 índices globais no período de 1.000 dias (07/01/2015 a 24/01/2019).

Índices	CCMP1	ASX11	CAC11	DAX11	HSI11	NYA11	SHCOM	SPTSH	TPX11	UKX11
Retorno Bruto (%)	**19,71**	**48,09**	**31,51**	**21,57**	**-9,31**	**32,27**	**122,88**	**10,94**	**-11,55**	**59,57**
Custo Operacional (%)	4,34	10,58	6,93	4,75	-2,05	7,10	27,03	2,41	-2,54	13,10
Retorno Líquido (%)	15,37	37,51	24,58	16,83	-7,26	25,17	95,85	8,54	-9,01	46,46
Buy And Hold (%)	**43,58**	**-7,67**	**9,81**	**6,41**	**4,31**	**9,61**	**-22,41**	**-7,76**	**18,38**	**-9,03**
Precisão (%)	**56,29**	**55,49**	**55,66**	**55,17**	**52,61**	**56,33**	**62,13**	**54,64**	**46,29**	**57,77**
Comprar Sempre (%)	19,78	14,32	-4,99	-13,07	-67,91	17,08	158,32	-27,24	-20,64	12,23
Inverso (%)	0,06	-33,77	-36,50	-34,65	-58,59	-15,19	35,44	-38,18	-9,09	-47,33
Aleatório	17,96	-13,53	-16,69	7,06	-19,18	-4,78	60,83	-19,78	25,06	22,56
Alvo 0,50%	9,62	-2,23	21,03	1,31	-7,70	24,27	62,07	4,96	-34,77	9,13
Alvo 1,00%	30,69	20,58	40,30	4,40	-5,21	33,92	88,07	5,37	-20,30	32,44
Alvo 2,00%	16,42	40,56	31,28	19,00	-0,49	35,19	107,24	10,91	-7,84	51,47
Perda (%)	-146,36	-141,30	-155,75	-161,22	-149,23	-89,96	-189,88	-105,12	-163,28	-149,73
Ganho (%)	166,08	189,39	187,26	182,79	139,91	122,23	312,76	116,06	151,73	209,30
Retorno Médio (%)	0,03	0,09	0,06	0,04	-0,02	0,07	0,21	0,02	-0,02	0,11
Perda Media (%)	-0,59	-0,59	-0,64	-0,69	-0,63	-0,42	-0,86	-0,48	-0,61	-0,65
Ganho Medio (%)	0,52	0,64	0,61	0,63	0,53	0,44	0,87	0,44	0,66	0,66
Acurácia (%)	51,05	52,75	54,65	53,75	54,65	53,35	54,95	54,25	49,05	55,76
F1score (%)	56,84	55,81	57,38	55,49	53,63	54,22	61,60	53,70	47,58	58,85
Revocação (%)	57,40	56,12	59,22	55,81	54,70	52,27	61,08	52,79	48,94	59,96
Especificidade (%)	42,92	48,93	49,79	51,55	54,62	54,56	46,08	55,73	49,15	51,06
Qtde. Perda	250	239	243	234	236	214	220	220	268	231
Qtde. Ganho	322	298	305	288	262	276	361	265	231	316
Qtde. Total	572	537	548	522	498	490	581	485	499	547
Qtde. Previsão Baixa	427	462	451	477	501	509	418	514	500	452
Qtde. Previsão Alta	572	537	548	522	498	490	581	485	499	547
Índice Sharpe (%)	-0,05	-0,02	-0,04	-0,02	-0,16	-0,02	0,52	-0,05	-38,00	0,00
Índice Sortino (%)	-0,07	-0,03	-0,06	-0,03	-0,24	-0,04	0,77	-0,07	-55,12	0,00
Média (%)	0,03	0,09	0,06	0,04	-0,02	0,07	0,21	0,02	-0,02	0,11
Desvio Padrão (%)	0,79	0,87	0,85	0,92	0,78	0,60	1,34	0,62	0,98	0,91
Desvio Absoluto (%)	0,54	0,61	0,62	0,66	0,58	0,43	0,84	0,45	0,63	0,65
Mediana (%)	0,09	0,12	0,08	0,07	0,02	0,06	0,22	0,07	0,00	0,15
Quantil (%)	0,09	0,12	0,08	0,07	0,02	0,06	0,22	0,07	0,00	0,15
Mínimo (%)	-4,24	-4,53	-3,38	-4,05	-3,62	-3,09	-6,52	-2,46	-7,90	-4,67
Máximo (%)	3,38	3,47	4,45	4,56	2,35	2,05	8,07	2,24	4,27	3,58
Amplitude (%)	7,62	8,00	7,83	8,61	5,97	5,14	14,59	4,70	12,17	8,25
Variancia (%)	0,62	0,76	0,73	0,85	0,61	0,36	1,79	0,39	0,97	0,84
Preco Inicial ($)	4790,53	5486,41	5149,97	12014,88	3028,26	10941,68	511,10	12597,85	11,92	10200,88
Preço Final ($)	7305,90	4979,45	5758,64	12942,55	3185,49	12208,06	373,12	11420,64	14,56	9091,18
Baixa (%)	43,80	46,80	48,40	48,30	52,00	47,10	40,90	49,70	52,80	47,30
Alta (%)	56,20	53,20	51,60	51,70	48,00	52,90	59,10	50,30	47,20	52,70
Qtde. Baixa	438	468	484	483	520	471	409	497	528	473
Qtde. Alta	562	532	516	517	480	529	591	503	472	527
Volatilidade	775,13	680,50	704,07	1807,97	429,74	1114,27	123,12	1037,33	1,86	1318,99
Volatilidade D.	530,82	430,70	438,95	1134,28	287,99	720,72	82,48	675,80	1,28	825,71
ShapiroWilk	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kolmogorov S.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Tstudent	0,73	0,10	0,16	0,27	0,27	0,15	0,46	0,14	0,96	0,04
Anova	0,42	0,00	0,00	0,16	0,20	0,00	0,00	0,03	0,42	0,00
kruskalWallis	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,78	0,00

Como os algoritmos de aprendizado de máquina utilizados nesta pesquisa são determinísticos, não foi necessário realizar um teste estatístico de hipótese com as técnicas tradicionais tais como: *Shapiro Wilk*, *Kolmogorov Smirnov*, *Tstudent*, *Anova*, e *kruskal Wallis*. No entanto, estas técnicas já estão contempladas no arcabouço e os valores utilizados no cálculo das mesmas, foram em relação aos retornos financeiros brutos e todos os dias em que

houve operação pelo modelo proposto, utilizando como referência a estratégia principal do Algoritmo 1.

Realizou-se o Teste de Hipótese *T Student* com a finalidade de verificar se o Retorno Financeiro Bruto da Estratégia de operação principal *RBM + NB* é estatisticamente diferente do Retorno Financeiro da Estratégia de operação *Comprar Sempre (Comprar S.* para os 10 índices globais avaliados no período amostral de 1.000 dias úteis. E somente foram encontrados valores para o *P-valor > 0,05* para os 2 índices globais (CCMP1 e TPX11), conforme destacado na coloração amarela na Tabela 15.

Já o Teste de Hipótese *KruskalWallis* foi realizado comparando as amostras de Retorno Financeiro Bruto (%) das 8 Estratégias de Operação Propostas (*RBM + NB*, Retorno Líquido (%) *Buy And Hold*, *Comprar Sempre*, *Inversa, Aleatória, Alvo 0,5%, Alvo 1,0% e Alvo 2,0%*). E somente foi encontrado valor para o *P-valor > 0,05* para os ínidce global TPX11, conforme destacado na Tabela supracitada. As medidas de risco Volatilidade, Volatilidade *Dowside*, índice *Sharpe* e índice *Sortino*, foram calculadas em relação ao retorno financeiro bruto diário de todos os dias em que houve operação pelo modelo proposto, utilizando como referência a estratégia principal do Algoritmo 1.

Para 7 índices (ASX11, CAC11, DAX11, NYA11, SHCOM, SPTSH e UKX11) o Retorno Financeiro Bruto encontrado pelo modelo *RBM + NB* foram superiores aos valores do *Buy and Hold*, conforme destacado na coloração verde, totalizando de acerto pelo modelo de classificação de 70,00% das previsões realizadas. Já a medida de Precisão deste modelo teve uma variação de 46,29% a 62,13%, com destaque para o índice SHCOM da China que apresentou o maior valor de Precisão e consequentemente o maior valor de Retorno Financeiro Bruto de 122,18%.

Nas Figuras 8 e 9 são apresentados séries de Preço de Fechamento e de Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB*. Para 6 dos índices globais (NYA11, CAC11, CCMP1, DAX11, HSI11 e TPX11) apresentam uma distribuição da série de preço de alta, ou seja, o valor do *Buy and Hold* foram positivos. Já os 4 índices (SHCOM, SPTH, ASX11 e UK11) as séries de preço de fechamento foram praticamente negativas. No entanto, ao avaliar a Tabela 15 praticamente os 10 índices globais apresentaram distribuições de Alta e baixa balanceadas contendo distancias entre as distribuições de no máximo 20%.

Figura 8 – Séries de Preço de Fehamento: 10 índices globais, 1.000 dias úteis (07/01/2015 a 24/01/2019).

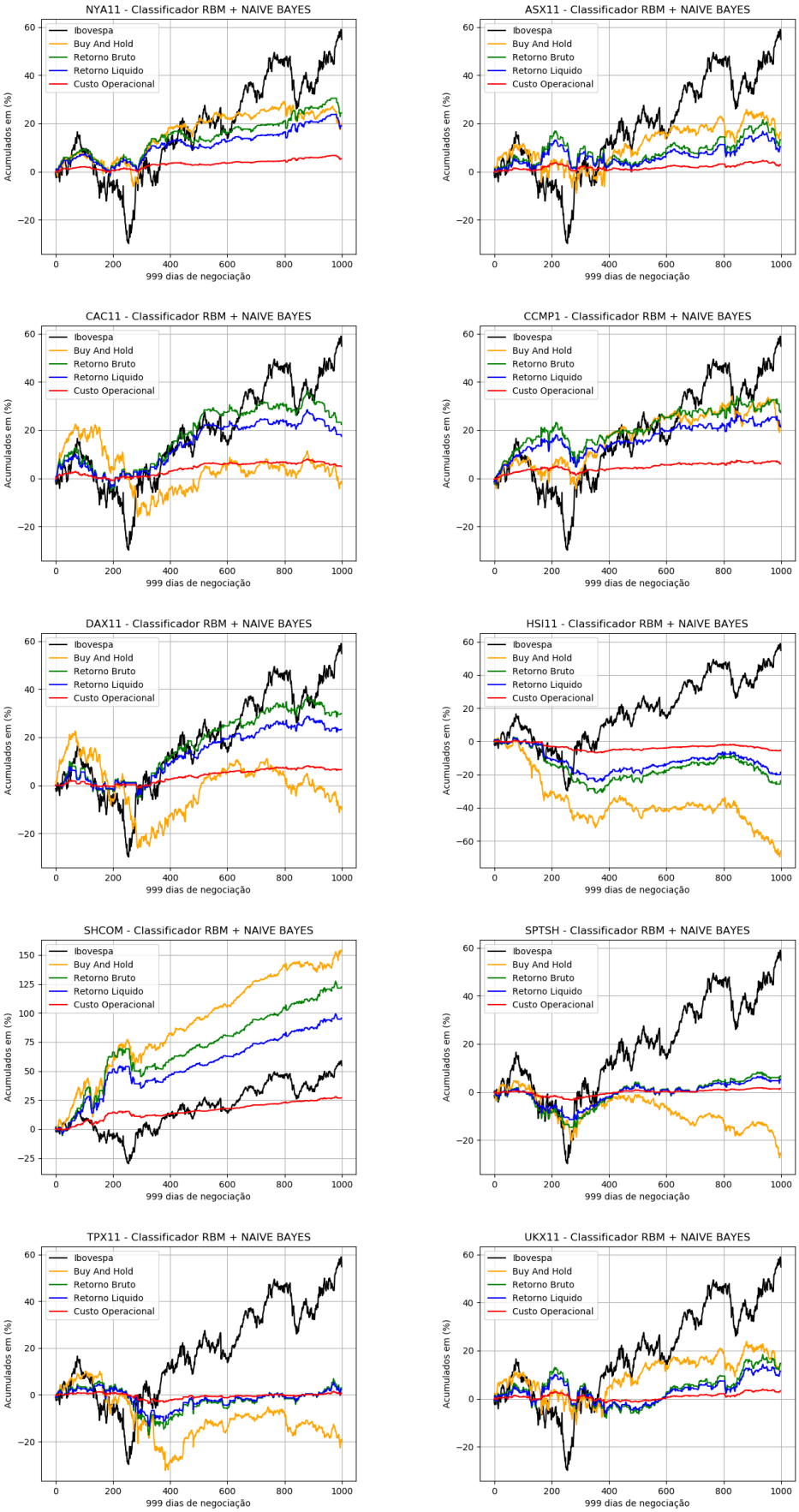


Figura 9 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 10 índices globais, 1.000 dias úteis (07/01/2015 a 24/01/2019).

5.3 Considerações Finais

Durante as simulações realizadas, conforme Tabela 16 para um período de 1.000 dias úteis na etapa de validação dos modelos de aprendizado de máquina propostos, o modelo combinado *RBM + NB* gerou como resultados da avaliação dos 50 ativos da B3, uma Precisão média de 50,45%, uma Acurácia média de 52,79% e um Retorno Financeiro Bruto médio de 68,29%, sendo 17,01% superior ao *baseline* de *Buy and Hold* médio de 56,67%.

Tabela 16 – validação: modelo *RBM + NB* - Resultado da média das simulações.

Mercados avaliados	**Ativos B3**	**Índices Globais**	**Média Geral**
Quantidade de Índices / Ativos	50	10	N/A
Período (dias úteis)	1.000	1.000	N/A
Amostra avaliada	50.000	10.000	N/A
Precisão Média	**50,45%**	**55,24%**	**52,85%**
Acurácia Média	52,79%	53,42%	53,11%
Retorno Financeiro Bruto Médio	**68,29%**	**33,57%**	**50,93%**
Buy and Hold Médio	56,67%	4,52%	30,60%
Ganho Médio acima do Buy and Hold	**17,01%**	**39,03%**	**28,02%**

Já nas simulações realizadas para os 10 índices globais para o mesmo período, o referido modelo encontrou uma Precisão média de 55,24%, uma Acurácia média de 53,42% e um Retorno Financeiro Bruto médio de 32,57%, sendo 39,03% superior ao *baseline* de *Buy and Hold* de ganho 4,52%.

Foram detalhadas neste Capítulo as configurações de parâmetros contendo vários experimentos em relação à calibração dos algoritmos, à escolha da janela de classificação a ser utilizada, à escolha do modelo de previsão de tendências mais adequado aos dados avaliados e, à verificação da viabilidade dos custos operacionais.

Capítulo 6

Conclusão

Nesta pesquisa foi desenvolvido um arcabouço metodológico para a modelagem, a implementação e a avaliação de estratégias de investimento utilizando, para isso, dados extraídos da Brasil Bolsa Balcão (B3), alguns conceitos de análise técnica e algoritmos de aprendizado de máquina para a geração de modelos de previsão de tendências aplicados ao mercado de ações.

Inicialmente foram investigados os principais trabalhos que abrangem o tema de caracterização, modelagem e previsão de séries temporais, para identificar conceitos e fundamentos, através da RSL. Verificou-se que o algoritmo *RBM* de categoria não supervisionada de *deep learning* está sendo muito empregado na comunidade acadêmica com a finalidade de realizar redução de dimensionalidade através da seleção e extração de características de dados de séries temporais.

Após contextualizar o estado da arte em relação aos principais trabalhos selecionados pela RSL, foram descritos alguns conceitos e características sobre a bolsa de valores de São Paulo (B3), o índice de mercado de referência (Ibovespa), a teoria da análise técnica, os custos operacionais praticados no mercado de ações, as aplicações dos algoritmos tradicionais de aprendizado de máquina, e as métricas de avaliação de desempenho, operação e risco, no Capitulo 3.

Implementou-se 4 modelos de previsão (*SVM, NB, AD J48* e *KNN*) baseados em algoritmos de aprendizado de máquina. Posteriormente foram realizadas mais 4 combinações com o algoritmo *RBM* de aprendizagem de categoria não supervisionada de (*deep learning*) contendo os algoritmos tradicionais citados anteriormente (*RBM + SVM, RBM + NB, RBM + AD J48* e *RBM + KNN*), a fim de realizar a seleção e extração de características dos atributos envolvidos, totalizando 8 propostas de modelos de previsão de tendências.

No Capítulo 5 foram apresentadas as distribuições das séries temporais do preço de fechamento original, a setorização e o *buy and hold* dos 50 ativos da B3 de setores distintos

e de 10 índices globais avaliados.

Várias simulações foram realizadas para a etapa de Treinamento do período de 1.000 dias úteis (2010 a 2014), contendo somente dados de cotações históricas de 50 ativos da B3, para realizar a configuração e calibragem de parâmetros dos 8 modelos de algoritmos de aprendizado de máquina propostos, validar as janelas de classificação, escolher o modelo de previsão de tendências e verificar a viabilidade dos custos operacionais. Como resultado, o modelo *RBM + NB* foi selecionado.

Com o objetivo de avaliar a robustez e capacidade de generalização dos 8 modelos de previsão de tendências propostos, para a etapa de Validação do período de 1.000 dias úteis (2015 a 2019), realizou-se mais duas simulações contendo dados de cotações históricas de 50 ativos da B3 mais dados de 10 índices globais e, para ambos os mercados, em relação ao indicador Retorno Financeiro Bruto, o melhor modelo de previsão de tendências novamente foi o *RBM + NB*.

Percebe-se nas simulações realizadas, que a variação do mercado dos 10 índices globais, em relação aos seus respectivos *Buy and Hold*, é bem discreta ao comparar com os 50 ativos do índice Bovespa da B3, ou seja, o mercado brasileiro de ações costuma ser mais volátil em relação aos demais mercados mundiais e, consequentemente está propício para estudos e aplicações de oportunidades e estratégias de negociação.

É importante ressaltar que o custo operacional não é tributado de forma proporcional, pois possui algumas taxas com valores fixos, além disso, o custo operacional é considerado muito elevado para aplicação em renda variável na bolsa de valores brasileira, dependendo do capital inicial disponível de pequenos e médio investidores, o que dificulta ainda o acesso e a popularização de investimentos neste seguimento de mercado.

Devido à existência de vários fatores externos de caráter político, econômico e social que interferem no mercado de ações, não é possível garantir que sempre que um modelo de previsão de tendências baseado em classificadores que obtenha uma precisão de desempenho acima de 50%, será encontrado também, um percentual de retorno financeiro líquido viável e acima dos *baselines* de mercado, o que pode inviabilizar toda uma proposta de estratégia de investimento.

Estes diversos intervalos associados a fatores externos como fatos e notícias (econômicas, políticas e sociais), provocam variações percentuais significativas, denominada *gap*, entre o preço de abertura e o preço de fechamento do dia anterior, os quais contribuem para inviabilizar muitas das estratégias de operação praticadas por agentes do mercado de ações.

6.1 Limitações e escopo

Durante os 5 anos de dedicação a esta tese de doutorado, foram pesquisados vários trabalhos acadêmicos, ferramentas de *backtest*, bases de dados, códigos implementados, técnicas de mercado e, em resumo, tive muita dificuldade em relação:

- **Bases de dados:** o acesso a grandes bases de dados (preço, volume, noticias, indicadores e etc) de forma online, tempo real, ainda é muito limitado, sendo necessário trabalhar na maioria das vezes com dados estáticos e de forma *offline*; e
- **Reutilização de códigos:** existiam muitos trabalhos na linha de pesquisa da minha tese, porém na hora de reproduzir ou aproveitar algo já existente na literatura, não existia acesso aos dados ou as técnica / código implementado disponibilizados estavam incompletos, sendo necessário implementar tudo do zero, o que demandou muito tempo para criar o arcabouço proposto.

Devido o arcabouço proposto implementado utilizar vários classificadores como previsores de tendências binárias, abre-se espaço para os seguintes estudos e avaliações:

- **Classes avaliadas:** nesta pesquisa avaliamos somente classes binarias (0 e 1), ou seja, pode ser um caminho avaliar as series temporais com multi classes;
- **Segmentação de previsores:** seria melhor um único algoritmo / técnica ou modelo de aprendizado de máquina para atuar em cada tipo de séries (Alta, Baixa ou Lado) de forma segmentada?
- **Novas fontes de dados:** avaliar o comportamento dos modelos de aprendizado de máquina com dados de criptomoedas, dados de mini índice e mini dólar;
- **Estratégias de risco:** neste trabalho não foi implementada estratégia baseada em indicadores e métricas de medidas de risco, nem isolada e nem de forma híbrida com os modelos de aprendizado de máquina;
- **Estratégias *ensemble*:** realizar a combinação de vários modelos de aprendizado de maquina para gerar uma saída global (ensemble);
- **Estratégias com *metaclassificadores:*** (ASSIS, 2019) para avaliar o comportamento destes novos modelos no arcabouço; e
- **Estratégias de mercado:** o arcabouço poderia ser ainda refinado acrescentando técnicas ou estratégias de mercado, como por exemplos, não realizar operações (compra e venda) entre 13:00 e 15:30, não operar nas segundas-feiras e sextas-feiras, devido fechamento em abertura do mês e vencimento de outros mercados.

Não foi escopo desta pesquisa trabalhar com algoritmos de aprendizado por reforço, aprofundar em avaliar vários algoritmos de *deep learning*, gestão e otimização de portfólio de carteiras de investimentos.

6.2 Publicações realizadas

Ao término do curso de Mestrado em Modelagem Matemática e Computacional em 2014, como aluno de disciplina isolada de doutorado em 2015 e durante 48 meses como aluno regular do curso de doutorado, realizei as seguintes publicações:

- (2020) **Proposta e Avaliação de Aplicações de Aprendizado de Máquina Utilizando Dados do Mercado Financeiro**, Revista Brasileira de Computação Aplicada (qualis B2), publicado na edição de abril de 2020;
- (2018) ***Proposal and Implementation of Machine Learning Models for Stock Markets using Web Data***, Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2018), no período de 16 a 19 de outubro de 2018, em Salvador (BA);
- (2018) ***Proposal and Implementation of Machine Learning and Deep Learning Models for Stock Markets***, *Workshop of Artificial Intelligence Applied to Finance (WAIAF)* em 14 de agosto de 2018 no Instituto Tecnológico de Aeronáutica (ITA);
- (2018) ***Hybrid Deep Learning Approach For Financial Time Series Classification***, Revista Brasileira de Computação Aplicada (RBCA) em agosto de 2018. Trabalho realizado em parceria com o Dr. Carlos Assis ex-aluno do MMC do CEFET-MG; e
- (2015) **Proposta e Implementação de Novas Estratégias de Negociação em Bolsa de Valores Usando Dados de Redes Sociais**, Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2015), no período de 27 e 30 de outubro de 2015, em Manaus (AM).

6.3 Trabalhos futuros

Atualmente estou envolvido nos seguintes trabalhos:

1. **Disponibilizar** o arcabouço em um servidor na Web para a utilização de forma gratuita para a sociedade acadêmica;
2. **Aplicar** os modelos de previsão de tendência em dados de criptomoedas;
3. **Implementar** estratégias de operação composta que avalie o risco X retorno utilizando os indicadores (*Sharpe* e *Sortino*) como filtro para o tratamento de possíveis ruídos;
4. **Implementar** outros modelos de tipos de aprendizado de máquina como por exemplo as redes LSTM e modelos de aprendizado por reforço; e
5. **Validar** o modelo *RBM + NB* proposto em um ambiente realístico que contemple todas as variáveis de operação do mercado de ações.

Referências

AAMODT, T. **Predicting stock markets with neural networks**. Dissertação (Mestrado) — University of Olso, 2015. Citado 2 vezes nas páginas 16 e 17.

ASSIS, C. **Predição de Tendências em séries financeiras utilizando Meta-classificadores**. Tese (Doutorado) — Centro Federal de Educação Tecnológica de Minas Gerais, 2019. Citado 3 vezes nas páginas 16, 17 e 60.

B3. O que a bolsa faz? **Brasil Bolsa Balcão**, 2014. Disponível em: <www.bmfbovespa.com.br>. Citado na página 20.

B3. Lista de corretoras da b3. acessado em 27 de dezembro de 2019 2019. Disponível em: <http://www.b3.com.br/pt_br/produtos-e-servicos/participantes/busca-de-participantes/busca-de-corretoras/>. Citado na página 20.

B3. Manual de procedimentos operacionais de negociação da b3. **Brasil Bolsa Balcão**, acessado em 28 de dezembro de 2019 2019. Disponível em: <www.bmfbovespa.com.br>. Citado na página 22.

B3. Histórico de pessoas físicas b3. acessado em 01 de janeiro de 2020 2020. Disponível em: <www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/historico-pessoas-fisicas/>. Citado na página 2.

B3-AçõES, B. B. B. Tarifas de ações e fundos de investimento. Fevereiro 2020. Disponível em: <http://www.b3.com.br/pt_br/produtos-e-servicos/tarifas/listados-a-vista-e-derivativos/renda-variavel/tarifas-de-acoes-e-fundos-de-investimento/a-vista/>. Citado na página 23.

B3-CUSTóDIA, B. B. B. Tarifas de serviços de custódia. Fevereiro 2020. Disponível em: <http://www.b3.com.br/pt_br/produtos-e-servicos/tarifas/servicos-da-central-depositaria/tarifas-de-servicos-de-custodia/>. Citado na página 23.

B3-VARIáVEL, B. B. B. Tarifas de renda variável. Fevereiro 2020. Disponível em: <http://www.b3.com.br/pt_br/produtos-e-servicos/tarifas/mercado-de-balcao/tarifas-de-renda-variavel/>. Citado na página 23.

BAUGHMAN, M. Predicting amazon spot prices with lstm networks. **Workshop on Scientific Cloud Computing**, 2018. Citado 2 vezes nas páginas 15 e 17.

BENGIO, Y. **Learning deep architectures for AI**. [S.l.]: Foundations and Trends in Machine Learning, 2009. Citado na página 24.

BOOTH, A. **Automated algorithmic trading: machine learning and agent-based modelling in complex adaptive financial markets**. Tese (Doutorado) — University of Southampton, 2016. Citado 2 vezes nas páginas 16 e 17.

CAO, W. Deep modeling complex couplings within financial markets. **Association for the Advancement of Artificial Intelligence**, 2015. Citado 3 vezes nas páginas 16, 17 e 43.

CETIP. **Cotações dos índices CDI e Ibovespa**. 2020. Disponível em: <http://bvmf.bmfbovespa.com.br/indices/ResumoVariacaoAnual.aspx?Indice=IBOV&idioma=pt-br>. Acesso em: 27 mar. 2020. Citado na página 1.

ERNST, A. Find the right people, processes and technology to manage record-to-report risks. **Managing Operational Tax Risk**, 2014. Citado na página 28.

ESTRADA, G. B. **Deep learning for multivariate financial time series**. Dissertação (Mestrado) — Royal Institute of Technology, 2015. Citado 3 vezes nas páginas 13, 16 e 17.

EXAME. **Bolsa já tem mais de 1,3 milhão de investidores**. 2019. Disponível em: <https://exame.abril.com.br/mercados/bolsa-ja-tem-mais-de-13-milhao-de-investidores/>. Acesso em: 27 jan. 2020. Citado na página 2.

EXAME. **Brasil vai dar adeus aos juros altos?** 2019. Disponível em: <https://exame.abril.com.br/economia/brasil-vai-dar-adeus-aos-juros-altos/>. Acesso em: 27 mar. 2020. Citado na página 1.

FERREIRA, J. C. P. **Análise de influência de utilizadores e redes sociais em microblogs sobre mercados financeiros**. Dissertação (Mestrado) — Engenharia e Gestão de Sistemas de Informação - Escola de Engenharia - Universidade do Minho, 2016. Citado 3 vezes nas páginas 12, 16 e 17.

GHOSH, A. et al. Stock price prediction using lstm on indian share market. In: YUAN, Q. et al. (Ed.). **Proceedings of 32nd International Conference on Computer Applications in Industry and Engineering**. EasyChair, 2019. (EPiC Series in Computing, v. 63), p. 101–110. ISSN 2398-7340. Disponível em: <https://easychair.org/publications/paper/LKgn>. Citado na página 17.

GIACOMEL, F. dos S. **Um método algorítmico para operações na bolsa de valores baseado em ensembles de redes neurais para modelar e prever os movimentos dos mercados de ações**. Dissertação (Mestrado) — Universidade federal do Rio Grande do Sul, 2016. Citado na página 14.

GOMIDE, J. S. A. **Mineração de redes sociais para detecção e previsão de eventos reais**. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Universidade Federal de Minas Gerais (UFMG), 2012. Disponível em: <www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/BUOS-8UKMG7/jana_nasant_annagomide.pdf?sequence=1>. Citado na página 12.

HAMILTON, W. P. **Stock market barometer: a study of its forecast value based on Charles H. Dow's theory of the price movement. With an analysis of the marketand its history since 1897**. [S.l.: s.n.], 1922. Citado na página 20.

HRASKO, R. Time series prediction using restricted boltzmann machines and backpropagation. **Information Technology and Quantitative Management**, 2015. Citado 4 vezes nas páginas 13, 16, 17 e 43.

ISHIKAWA, A. M. F. E. S. **Mercado Financeiro e de Capitais**. [S.l.: s.n.], 2000. Citado na página 20.

JIA, H. Investigation into the effectiveness of long short term memory networks for stock price prediction. **Cornell University Library**, 2016. Citado 4 vezes nas páginas 14, 16, 17 e 43.

KIMURA, H. Machine learning models and bankruptcy prediction. **Expert Systems With Applications**, v. 83, p. 405–417, 2017. Citado 4 vezes nas páginas 15, 16, 17 e 43.

KITCHENHAM, B. Guidelines for performing systematic literature reviews in software engineering. 2007. Citado 2 vezes nas páginas 6 e 7.

KLABJAN, D. Classification-based financial markets prediction using deep neural networks. 2016. Citado 3 vezes nas páginas 16, 17 e 43.

LANGONI, C. G. **História contada do Banco Central do Brasil**. [S.l.: s.n.], 2019. Citado na página 1.

LEE, J. Deep learning-based corporate performance prediction model considering technical capability. **Substainability**, 2017. Citado 4 vezes nas páginas 15, 16, 17 e 43.

LIU, J. Restricted boltzmann machine based stock market trend prediction. **International Joint Conference on Neural Networks**, 2017. Citado 4 vezes nas páginas 15, 16, 17 e 43.

LIU, T. Deep learning for event-driven stock prediction. **International Joint Conference on Artificial Intelligence**, 2015. Citado 4 vezes nas páginas 13, 16, 17 e 43.

LIUY, T. Deep learning for event-driven stock prediction. **Proceedings of the twenty-fourth international Joint Conference on Artificial Intelligence**, 2015. Citado na página 43.

MANSOURI, E. A. **Using artificial neural networks and sentiment analysis to predict upward movements in stock price**. Dissertação (Mestrado) — Worcester Polytechnic Institute, 2016. Citado 2 vezes nas páginas 16 e 17.

MARQUES, W. L. **Gerenciamento Financeiro**. [S.l.: s.n.], 2013. Citado na página 19.

MIRANDA, A. W. M. **Estratégias de Investimentos baseadas em microestrutura de mercado**. Dissertação (Mestrado) — Departamento de Ciência da Computação do Universidade Federal de Minas Gerais, 2019. Citado 2 vezes nas páginas 16 e 17.

NAMETALA, C. A. L. **Construção de um robô investidor baseado em redes neurais artificiais e preditores econométricos**. Dissertação (Mestrado) — Escola de Engenharia da Universidade Federal de Minas Gerais (UFMG), 2017. Citado 3 vezes nas páginas 14, 16 e 17.

PEARL, J. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988. ISBN 0934613737. Citado na página 25.

PINHEIRO, J. L. **Mercado de Capitais: fundamentos e técnicas**. [S.l.: s.n.], 2019. 616 p. ISBN 8597021438, 9788597021431. Citado na página 19.

QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 1558602380. Citado na página 24.

RAMAKRISHNAN, R. When is nearest neighbor meaningful? **CiteSeerX**, 1999. Citado na página 26.

RAMALHO, I. **Redes neurais artificiais na predição das principais séries do Indice Ibovespa e suas Aplicações em Sistemas Automatizados de Negociação**. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, 2014. Citado 2 vezes nas páginas 16 e 17.

ROCA, A. L. R. M. Renata Laíse Reis de S. Previsão do Índice bovespa por meio de redes neurais artificiais. **Business and Management Review**, 2015. Citado 3 vezes nas páginas 13, 16 e 17.

SALAKHUTDINOV, H. Reducing the dimensionality of data with neural networks. **Science**, v. 313, n. 5786, p. 504–507, 2006. Citado na página 26.

SANTOS, H. S. **Um estudo sobre o mercado brasileiro de ações a partir de dados do twitter**. Dissertação (Mestrado) — Mestrado em Ciências da Computação da Universidade Federal de Minas Gerais, 2016. Citado 3 vezes nas páginas 12, 16 e 17.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning : from theory to algorithms**. [s.n.], 2014. ISBN 9781107057135 1107057132. Disponível em: <http://www.worldcat.org/search?qt=worldcat_org_all&q=9781107057135>. Citado na página 23.

SHARANG, A. Using machine learning for medium frequency derivative portfolio trading. **Cornell University Library**, 2015. Citado 3 vezes nas páginas 16, 17 e 43.

SILVA, D. N. E. da. **Predição de tendência de ativos em séries temporais financeiras utilizando algoritmos de aprendizado de máquina**. Dissertação (Mestrado) — Instituto Militar de Engenharia, 2014. Citado 3 vezes nas páginas 12, 16 e 17.

SIMON, H. **Neural networks: A comprehensive foundation**. 2nd. ed. [S.l.]: Prentice Hall PTR, 1998. ISBN 0132733501. Citado na página 27.

VAPNIK, V. N. **The nature of statistical learning theory**. [S.l.]: Springer, 1995. Citado na página 25.

VARGA, G. índice de sharpe e outros indicadores de performance aplicados a fundos de ações brasileiros. **Revista de Administração Contemporanea**, 2001. Citado na página 28.

VENGERTSEV, D. **Deep learning architecture for univariate time series forecasting**. Dissertação (Mestrado) — Department of Computer Science Stanford University, 2014. Citado 3 vezes nas páginas 12, 16 e 17.

XIONG, R. Deep learning stock volatility with google domestic trends. **Cornell University Library**, 2016. Citado 4 vezes nas páginas 14, 16, 17 e 43.

ZHANG, Y. **Using financial reports to predict stock market trends with machine learning techniques**. Dissertação (Mestrado) — University of Oxford, 2015. Citado 3 vezes nas páginas 16, 17 e 27.

Apêndices

APÊNDICE A – 50 ativos da B3: Gráficos de séries de Preço de Fechamento

Nas Figuras 10, 11, 12, 13 e 14 são apresentadas as séries de Preço de Fechamento para o período de 1.000 dias úteis (07/01/2015 a 24/01/2019), de um total de 50 ativos que foram extraídos da Bolsa Brasil Balcão (B3).

Para 15 ativos (CIEL3, BRFS3, MULT3, LREN3, QUAL3, WEGE3, GOAU4, EMBR3, BRML3, RENT3, CCRO3, TOTS3, BBSE3, PCAR4 e ECOR3) o *buy and hold* foi negativo, variando de 0,16% a -61,37, totalizando aproximadamente 29,00% da amostra.

Já para os demais 37 ativos (VIVT4, MRFG3, TIMP3, UGPA3, BBDC3, ITUB4, CMIG4, LAME4, BBDC4, VVAR3, JBSS3, FLRY3, ITSA4, HGTX3, CSAN3, GOLL4, CYRE3, IGTA3, GGBR4, ENBR3, MRVE3, CSNA3, BTOW3, HYPE3, USIM5, BBAS3, RADL3, SBSP3, BRAP4, VALE3, PETR4, BRKM5, PETR3, CVCB3, ELET6, ELET3 e MGLU3) o *buy and hold* foi positivo, variando de 0,95% a 1.703,21%, totalizando aproximadamente 71,00% da amostra.

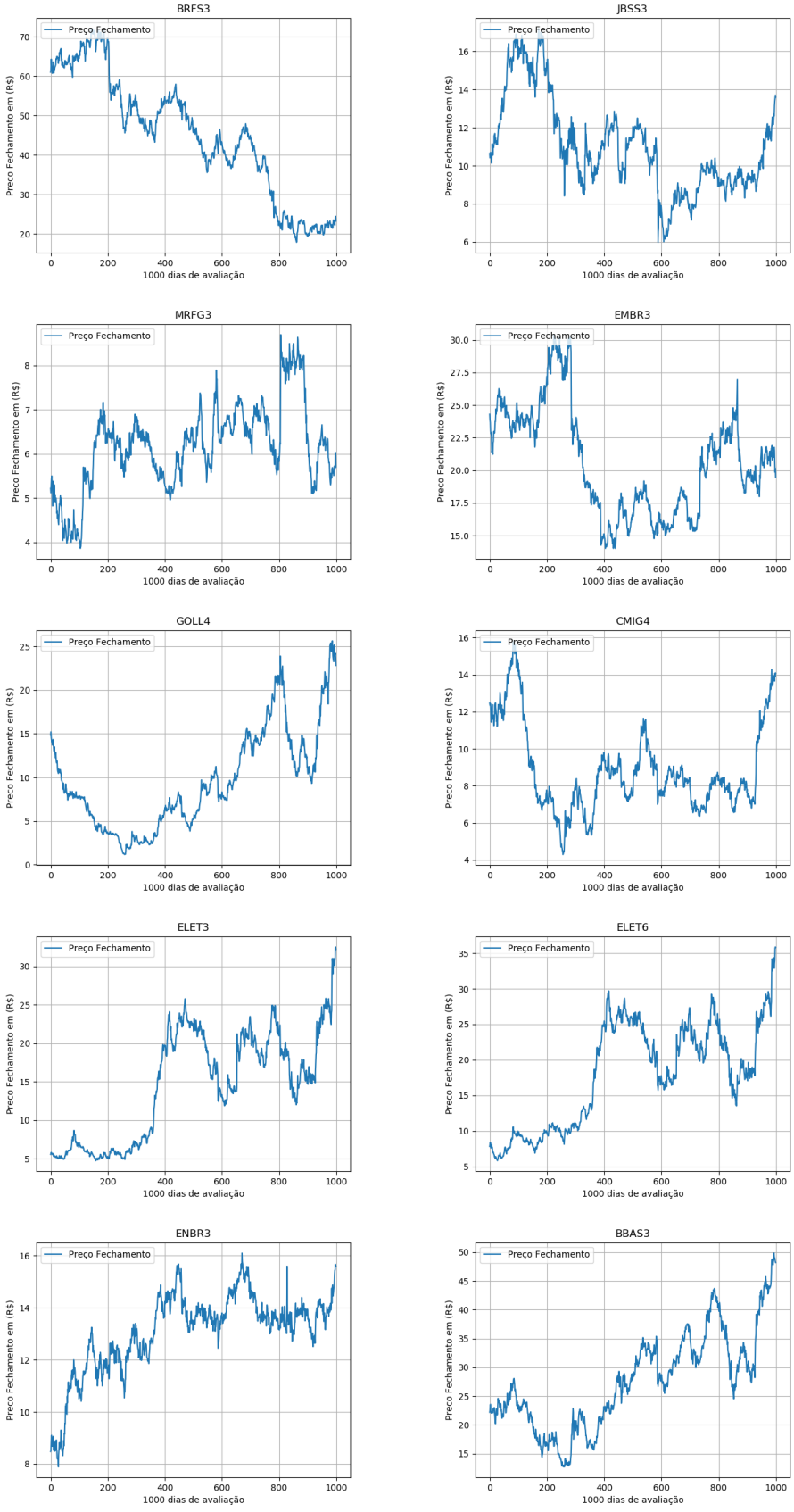


Figura 10 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 1 de 5

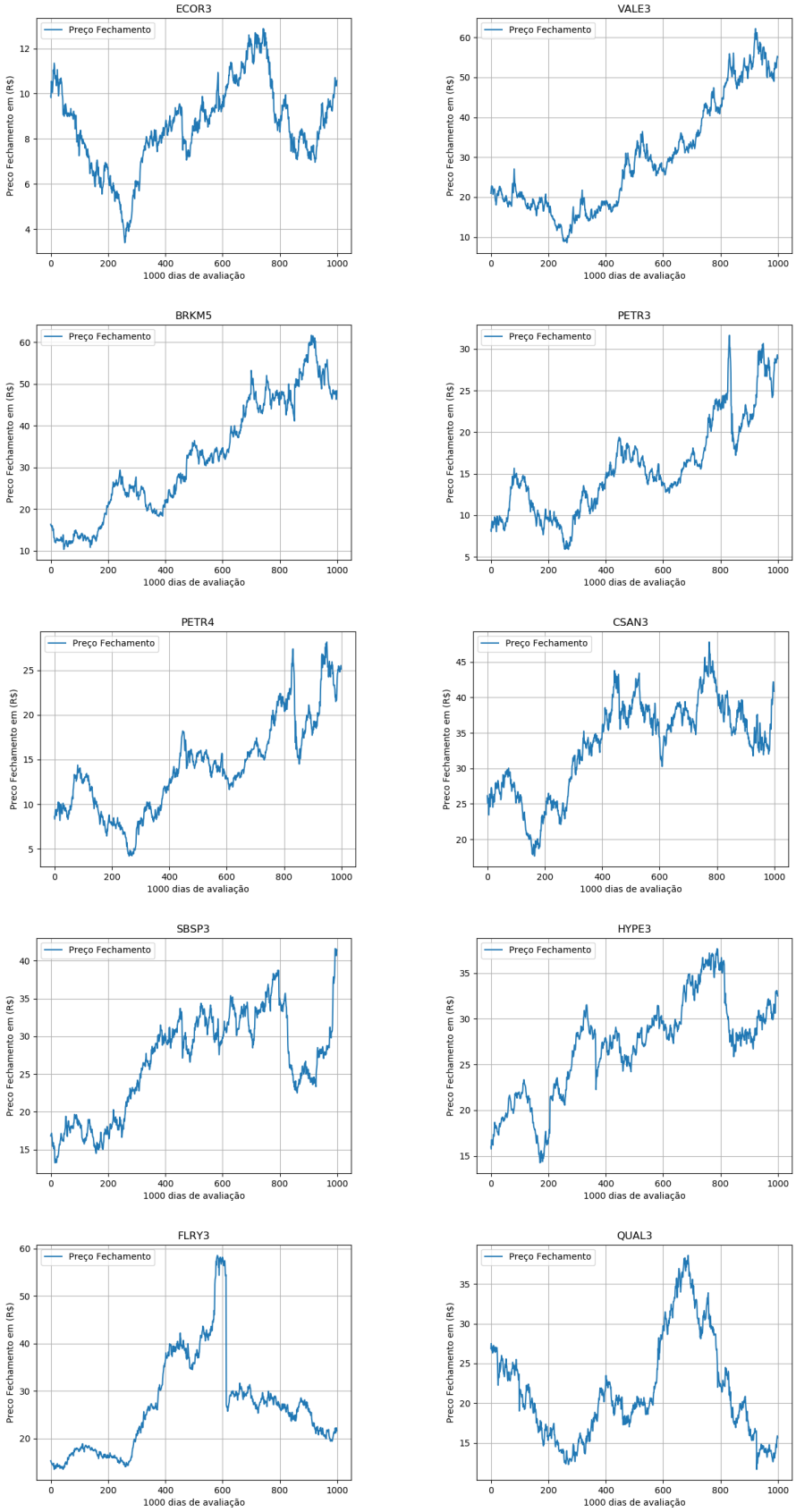


Figura 11 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 2 de 5

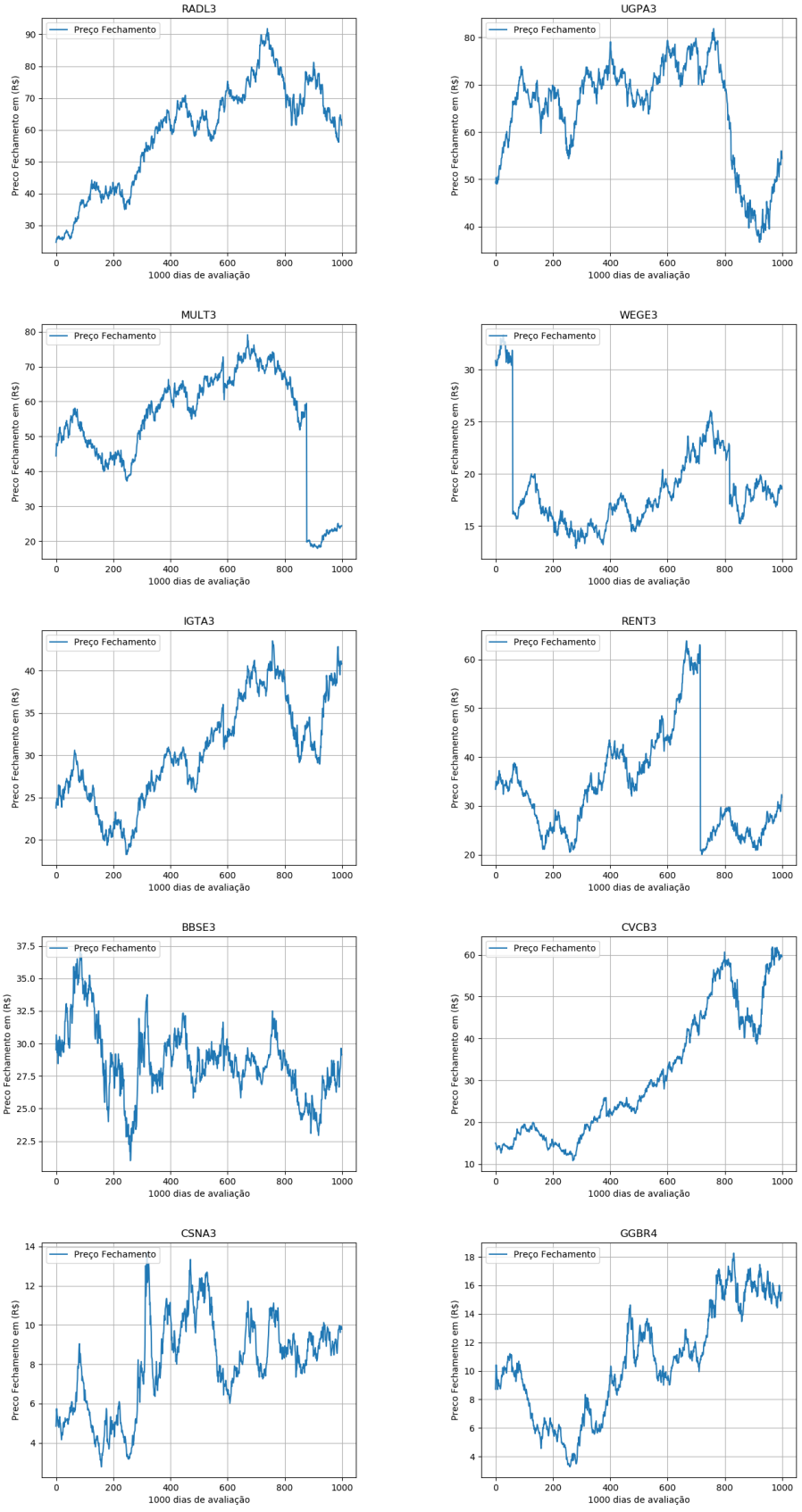


Figura 12 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 3 de 5.

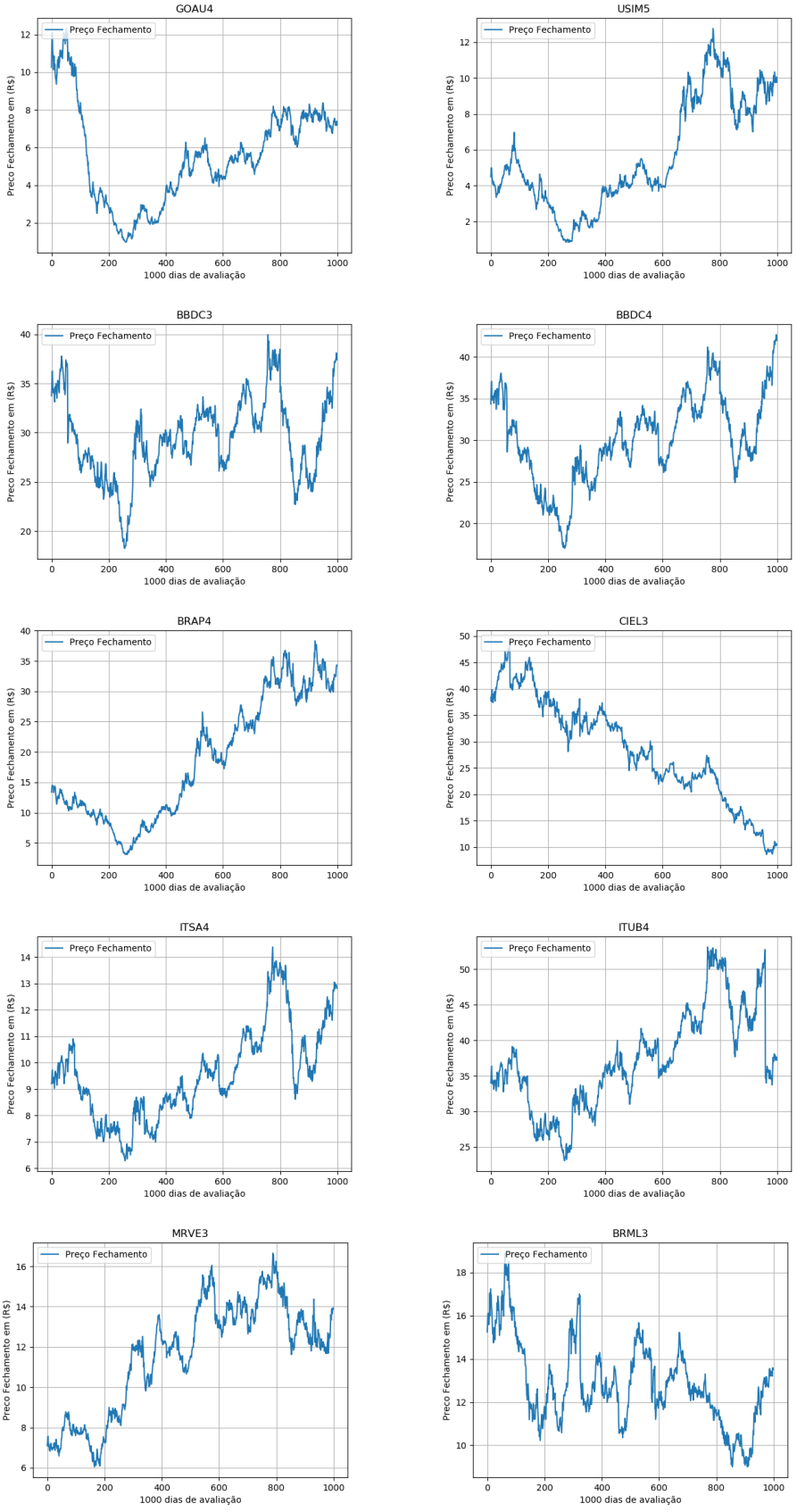


Figura 13 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 4 de 5.

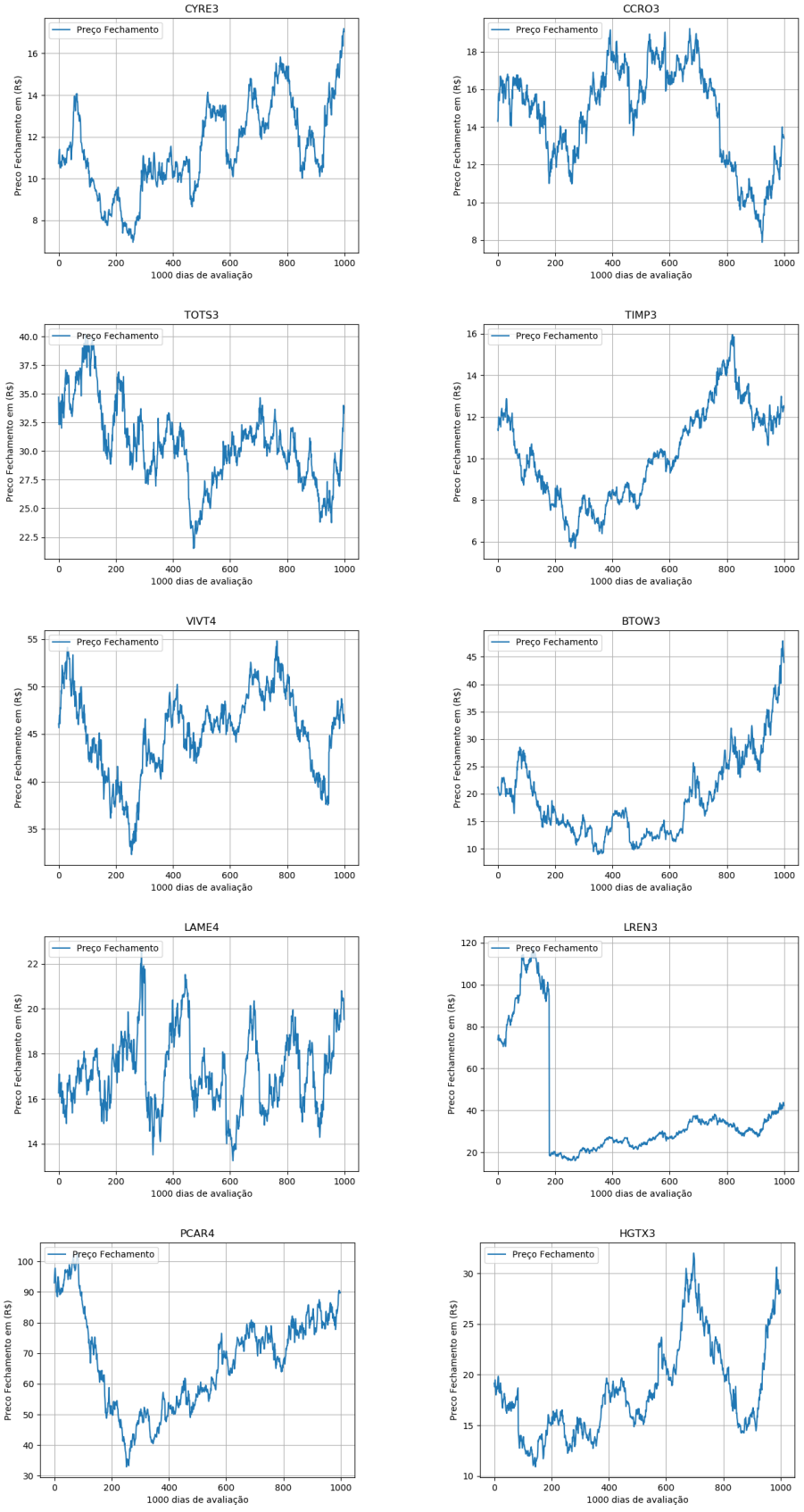


Figura 14 – Séries de Preço de Fehamento: 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 5 de 5.

APÊNDICE B – 50 ativos da B3: Gráficos de Retornos Acumulados

Nas Figuras 15, 16, 17, 18 e 19 são apresentadas as séries de Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* dos 50 ativos avaliados para o período de 1.000 dias úteis (07/01/2015 a 24/01/2019).

Para um total de 50 ativos da B3 avaliados, somente 15 ativos (GOLL4, CMIG4, ELET3, ELET6, BRAP4, ITSA4, VALE3, PETR3, USIM5, PETR4, SBSP3, RADL3, CVCB3, CSNA3 e GGBR4), como mostra a série de Retorno Financeiro Bruto em (%) na coloração em verde encontrado pelo modelo *RBM + NB*. obtiveram valores inferiores aos seus respectivos *Buy and Hold*, conforme destacado na coloração verde, totalizando um acerto pelo modelo de classificação proposto de 74,00% das previsões realizadas.

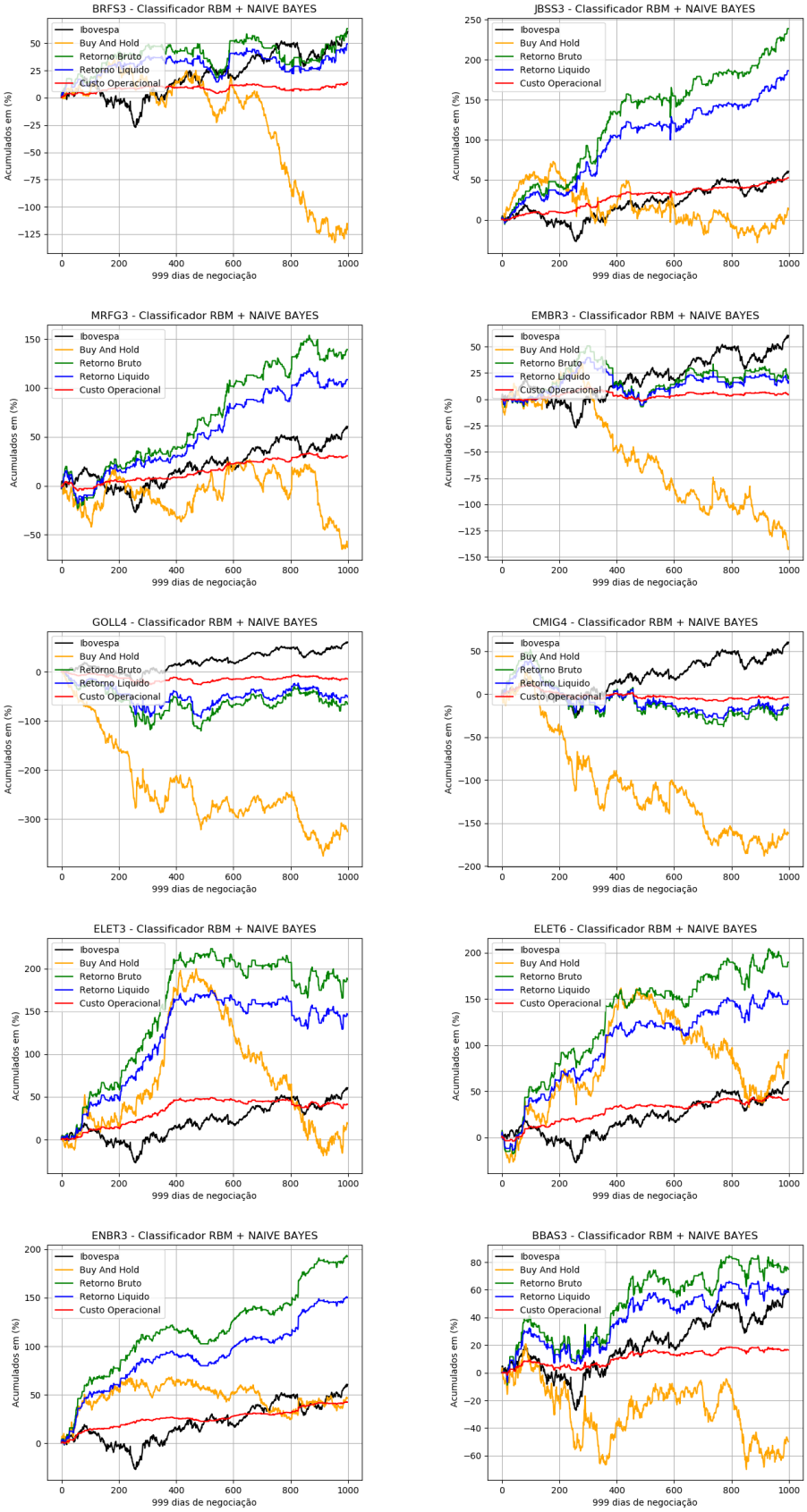


Figura 15 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 1 de 5.

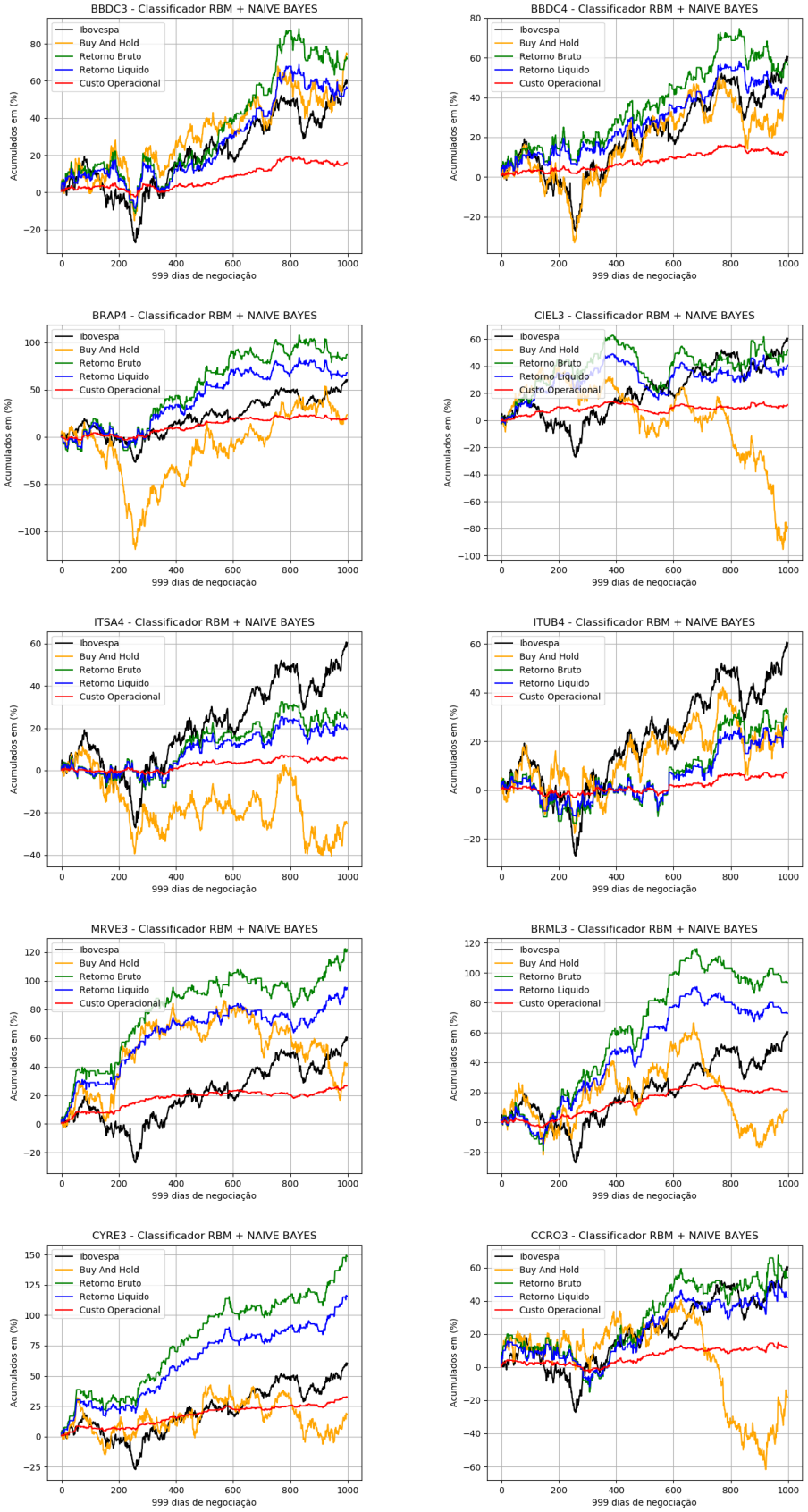


Figura 16 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 2 de 5.

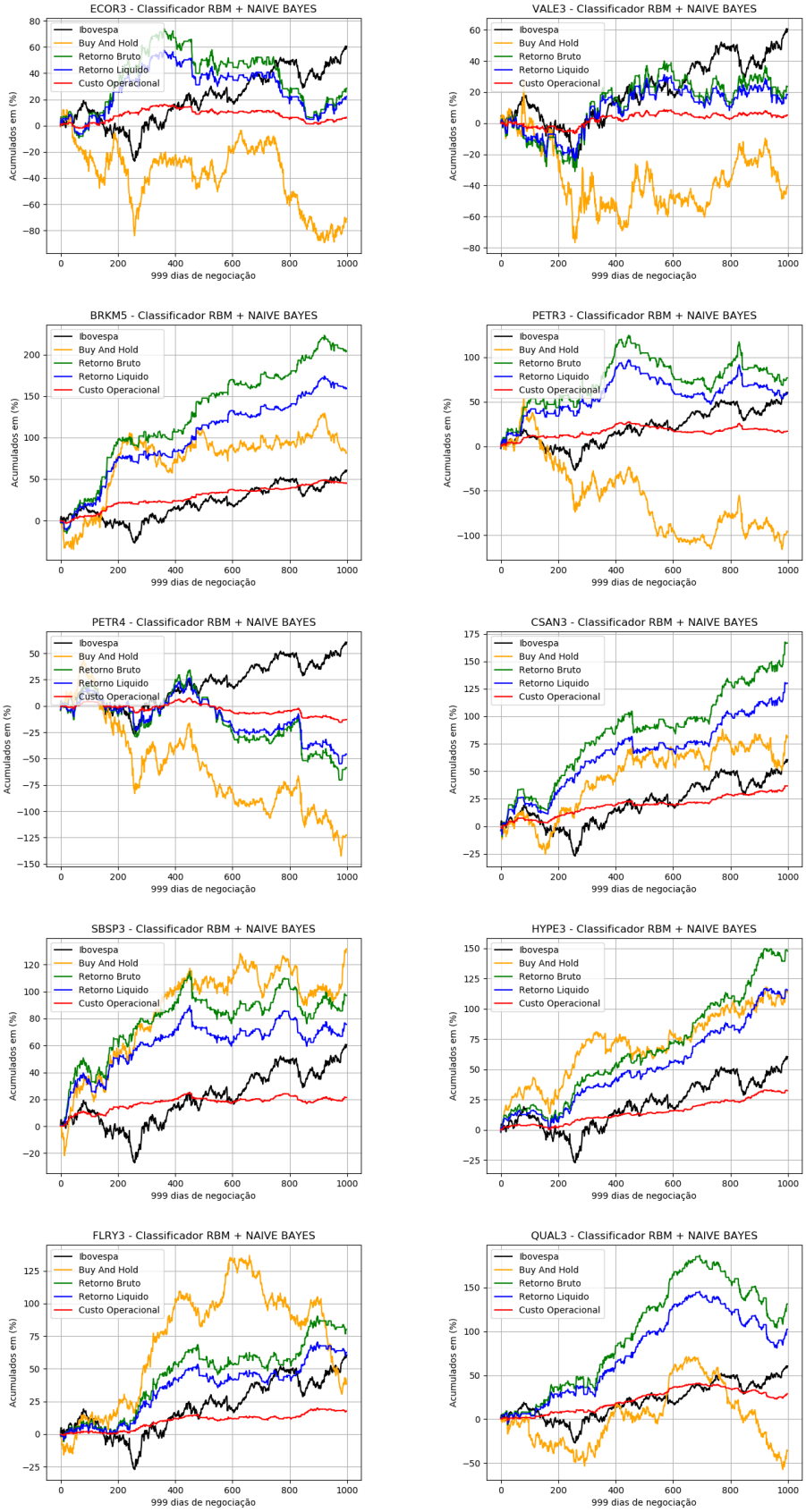


Figura 17 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 3 de 5.

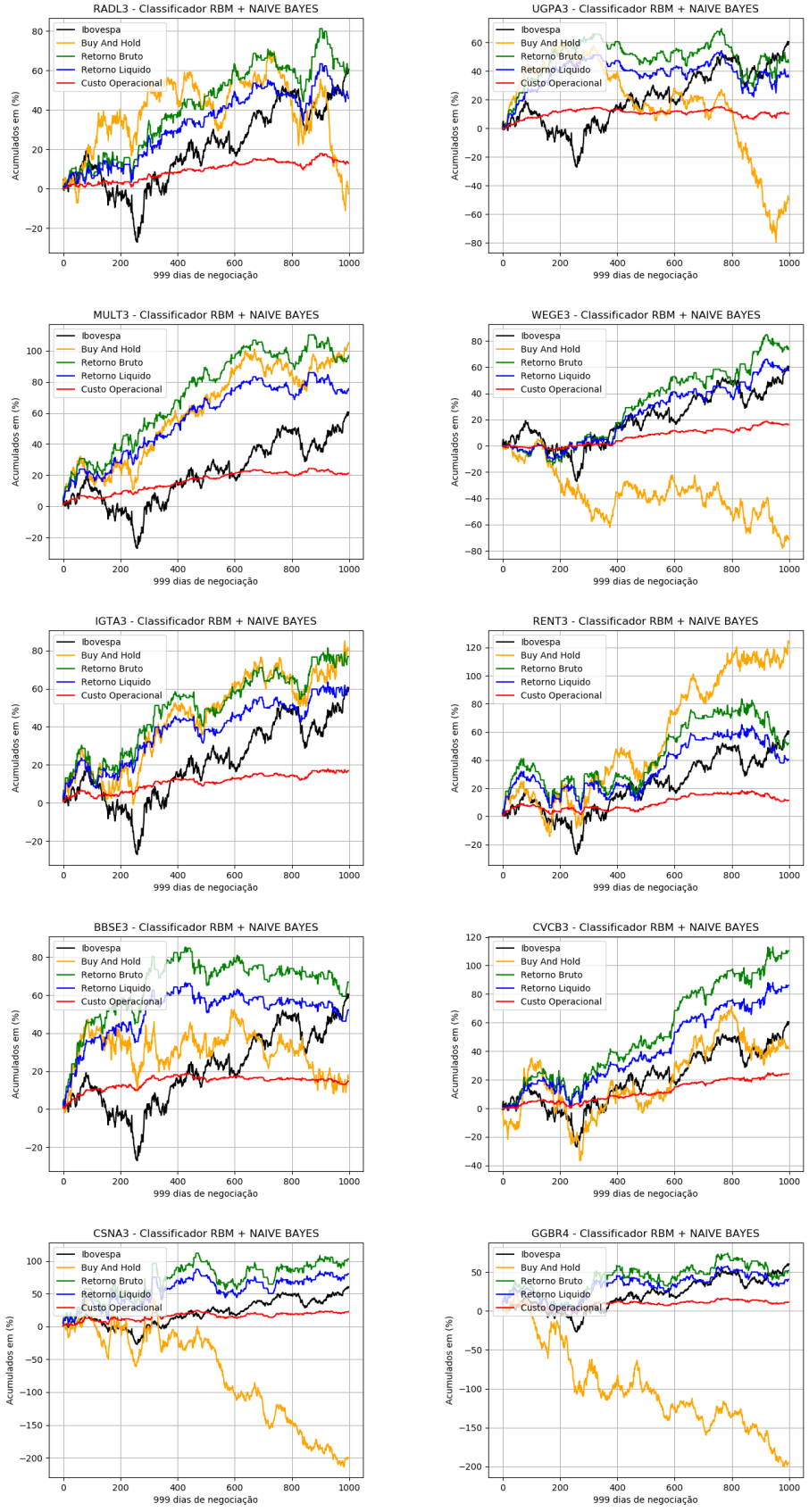


Figura 18 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 4 de 5.

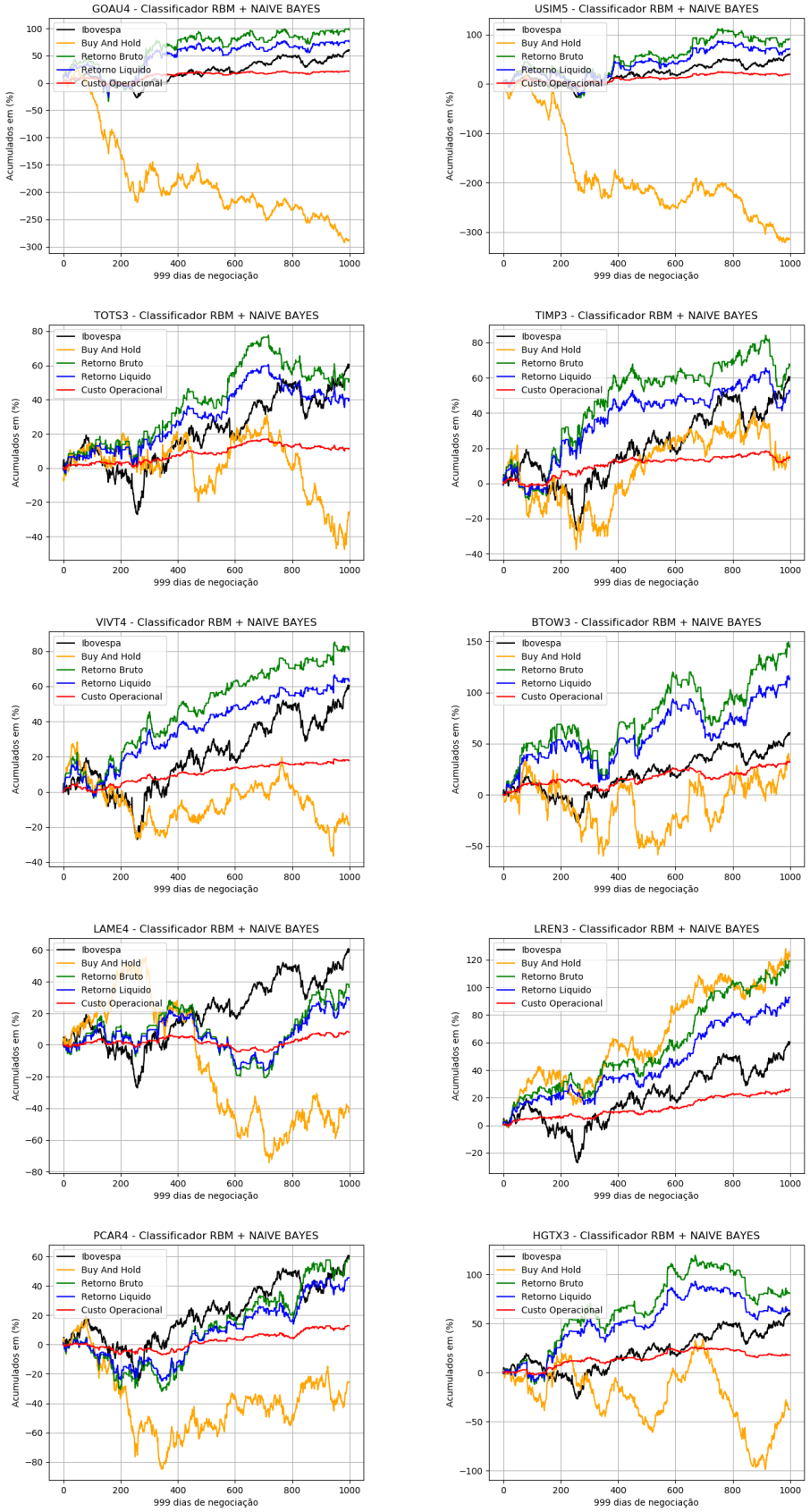


Figura 19 – Retornos financeiros (Bruto e Líquido), custo operacional, *buy and hold* e Ibovepa em (%) acumulados da simulação do modelo *RBM + NB* para 50 ativos, 1.000 dias úteis (07/01/2015 a 24/01/2019 - Parte 5 de 5.

APÊNDICE C – 50 ativos da B3: Resultados detalhados

Nas Tabelas 17, 18, 19, 20 e 21 são apresentados os resultados da etapa de Validação dos 50 ativos avaliados do modelo de previsão de tendências selecionado *RBM + NB*, contendo o extrato de todos os indicadores de análise estatística, desempenho e operação / risco que são gerados automaticamente pelo arcabouço proposto implementado, conforme já detalhados no Capítulo 4, nas descrições dos Algoritmos 1, 2, 3, 4, 5 e 6.

Para um total de 50 ativos da B3 avaliados, somente 16 ativos o indicador Retorno Financeiro Bruto encontrado pelo modelo *RBM + NB* obtiveram valores inferiores aos seus respectivos *Buy and Hold*, conforme destacado na coloração verde, totalizando um acerto pelo modelo de classificação de 74,00% das previsões realizadas. Já a medida de Precisão deste modelo teve uma variação de 44,19% a 54,56%, com destaque para o ativo ENBR3 (Energias do Brasil) que apresentou o maior valor de Precisão.

Em destaque na coloração em amarelo, estão os valores abaixo de 50,00% encontrados nas medidas de desempenho (Acurácia e Precisão), o que não significa que estes valores encontrados são ruins em se tratando de utilizar modelos de classificação para a predição de séries temporias de mercado de ações, uma vez que ainda sim, geraram Retornos Financeiros Brutos positivos e acima dos respectivos *Buy and Hold*, como foi o caso de 11 ativos (MRFG3, EMBR3, BBDC3, BBDC4, ECOR3, BRKM5, WEGE3, BBSE3, GOAU4, TOTS3 e PCAR4).

Realizou-se o Teste de Hipótese *T Student* com a finalidade de verficar se o Retorno Financeiro Bruto da Estratégia de operação principal *RBM + NB* é estatisticamente diferente do Retorno Financeiro da Estratégia de operação *Comprar Sempre* para os 50 ativos avaliados no período amostral de 1.000 dias úteis. E somente foram encontrados valores para o *P-valor > 0,05* para OS 9 ativos (CMIG4, BRAP4, ITSA4, ITUB4, VALE3, PETR4, MULT3, PCAR4 e LREN3), conforme destacado na coloração amarela nas Tabelas 17, 18, 19, 20 e 21.

Já o Teste de Hipótese *ANOVA* foi realizado comparando as amostras de Retorno Financeiro Bruto (%) das 8 Estratégias de Operação Propostas (*RBM + NB*, Retorno Líquido (%) *Buy And Hold*, *Comprar Sempre*, *Inversa, Aleatória, Alvo 0,5%, Alvo 1,0% e Alvo 2,0%*). E somente foram encontrados valores para o *P-valor > 0,05* para os 5 ativos (BRAP4, ITSA4, ECOR3, GGBR4 e PCAR4). conforme destacado nas Tabelas supracitadas.

Tabela 17 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 1 de 5.

Ativos	BRFS3	JBSS3	MRFG3	EMBR3	GOLL4	CMIG4	ELET3	ELET6	ENBR3	BBAS3
Retorno Bruto (%)	**70,10**	**166,17**	**104,94**	**28,20**	**-17,99**	**-28,52**	**140,73**	**192,52**	**161,20**	**40,19**
Custo Operacional (%)	15,42	36,56	23,09	6,20	-3,96	-6,28	30,96	42,35	35,46	8,84
Retorno Líquido (%)	54,68	129,61	81,85	22,00	-14,03	-22,25	109,77	150,16	125,74	31,35
Buy And Hold (%)	**-50,51**	**31,27**	**10,87**	**-14,80**	**45,36**	**13,21**	**421,53**	**295,16**	**77,56**	**94,11**
Comprar Sempre (%)	-122,21	14,14	-60,65	-139,53	-323,54	-157,89	18,73	88,86	46,94	-52,36
Inversa (%)	-192,31	-152,03	-165,60	-167,73	-305,56	-129,37	-122,01	-103,66	-114,26	-92,54
Aleatória (%)	-68,24	37,98	-4,37	-127,80	-217,06	73,68	91,54	-3,08	49,38	31,96
Alvo 0,5%	-14,05	7,38	37,14	-13,36	39,80	-40,17	-11,81	16,35	73,45	19,03
Alvo 1,0%	16,33	16,77	66,46	17,10	38,70	-42,48	5,91	64,67	115,10	-35,81
Alvo 2,0%	32,85	23,71	41,65	8,13	-18,29	-49,03	41,27	130,56	139,48	-23,05
Perda Total (%)	-318,27	-476,66	-409,25	-381,34	-710,85	-515,78	-541,69	-481,82	-273,49	-450,58
Ganho Total (%)	388,37	642,83	514,20	409,54	692,86	487,26	682,42	674,34	434,69	490,77
Retorno Médio (%)	0,15	0,33	0,22	0,06	-0,04	-0,06	0,29	0,39	0,31	0,08
Perda Média (%)	-1,35	-1,96	-1,68	-1,51	-2,69	-2,19	-2,28	-1,99	-1,17	-1,92
Ganho Médio (%)	1,61	2,49	2,29	1,73	3,32	2,24	2,80	2,69	1,55	1,90
Precisão (%)	**50,52**	**51,50**	**47,97**	**48,37**	**44,19**	**48,12**	**50,62**	**50,91**	**54,56**	**52,33**
Acurácia (%)	**55,66**	**54,55**	**54,95**	**52,75**	**52,55**	**53,15**	**53,55**	**53,85**	**55,96**	**54,35**
F1 Score (%)	52,11	53,20	50,00	50,11	46,86	48,23	51,26	52,13	56,09	53,09
Revocação (%)	53,79	55,01	52,20	51,97	49,88	48,34	51,91	53,40	57,70	53,86
Especificidade (%)	57,17	54,15	57,04	53,41	54,48	57,12	55,01	54,25	54,30	54,81
Qtde Perda	236	243	244	253	264	235	238	242	234	235
Qtde Ganho	241	258	225	237	209	218	244	251	281	258
Qtde Total	477	501	469	490	473	453	482	493	515	493
Qtde Previsão Baixa	522	498	530	509	526	546	517	506	484	506
Qtde Previsão Alta	477	501	469	490	473	453	482	493	515	493
Índice *Sharpe* (%)	0,15	19,12	14,91	-5,48	-13,63	-24,43	8,70	13,52	25,24	-2,77
Índice *Sortino* (%)	0,26	31,92	25,40	-8,69	-21,94	-39,95	14,31	23,34	44,88	-4,69
Média (%)	0,15	0,33	0,22	0,06	-0,04	-0,06	0,29	0,39	0,31	0,08
Desvio Padrão (%)	1,99	3,36	2,69	2,42	4,38	2,99	3,42	3,06	1,91	2,56
Desvio Absoluto (%)	1,48	2,24	1,98	1,62	2,96	2,21	2,54	2,35	1,37	1,91
Mediana (%)	0,05	0,14	0,00	0,00	-0,34	-0,14	0,16	0,09	0,18	0,12
Quantil (%)	0,05	0,14	0,00	0,00	-0,34	-0,14	0,16	0,09	0,18	0,12
DrawDown (%)	-4,64	-25,71	-15,45	-13,69	-22,17	-18,60	-12,95	-8,19	-5,54	-14,68
Máximo (%)	9,14	25,73	12,82	22,50	35,71	13,43	20,42	14,97	15,56	8,16
Amplitude (%)	13,78	51,44	28,27	36,19	57,88	32,03	33,37	23,16	21,10	22,84
Variância (%)	3,96	11,30	7,24	5,84	19,21	8,95	11,66	9,34	3,65	6,57
Preço Inicial (R$)	61,80	10,51	5,19	23,94	15,15	12,50	5,59	8,11	8,55	22,94
Preço Final (R$)	24,19	14,47	5,87	19,67	23,43	14,49	33,98	36,95	16,54	48,95
Distribuição Classe 0 (%)	55,10	53,00	56,90	54,40	58,00	54,90	52,90	52,90	51,20	52,00
Distribuição Classe 1 (%)	44,90	47,00	43,10	45,60	42,00	45,10	47,10	47,10	48,80	48,00
Qtde Real de Baixa	551	530	569	544	580	549	529	529	512	520
Qtde Real de Alta	449	470	431	456	420	451	471	471	488	480
Volatilidade (%)	12,86	5,13	2,47	7,28	6,32	3,96	8,35	9,20	3,69	10,08
Volatilidade *Dowside* (%)	7,68	3,07	1,45	4,60	3,92	2,42	5,08	5,33	2,07	5,96
Shapiro Wilk	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kolmogorov Smirnov	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
T. student	0,01	0,07	0,05	0,11	0,23	0,58	0,14	0,07	0,01	0,33
Anova	0,00	0,00	0,00	0,01	0,01	0,20	0,01	0,00	0,00	0,20
Kruskal Wallis	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Tabela 18 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 2 de 5.

Ativos	BBDC3	BBDC4	BRAP4	CIEL3	ITSA4	ITUB4	MRVE3	BRML3	CYRE3	CCRO3
Retorno Bruto (%)	**80,63**	**41,85**	**-4,56**	**10,24**	**-8,83**	**27,51**	**126,06**	**82,09**	**80,74**	**82,45**
Custo Operacional (%)	17,74	9,21	-1,00	2,25	-1,94	6,05	27,73	18,06	17,76	18,14
Retorno Líquido (%)	62,89	32,65	-3,56	7,99	-6,88	21,46	98,33	64,03	62,98	64,31
Buy And Hold (%)	**9,04**	**16,65**	**122,88**	**-59,57**	**32,94**	**5,90**	**84,69**	**-8,65**	**46,97**	**-1,65**
Comprar Sempre (%)	72,26	40,49	19,43	-76,03	-24,91	26,69	43,83	10,24	14,64	-23,10
Inversa (%)	-8,37	-1,37	24,00	-86,27	-16,09	-0,82	-82,23	-71,85	-66,11	-105,55
Aleatória (%)	48,93	25,02	-84,96	-38,81	-61,83	-13,84	21,65	-44,66	63,97	-13,73
Alvo 0,5%	44,23	22,33	-91,38	-22,97	5,82	40,76	43,84	52,36	59,56	10,62
Alvo 1,0%	94,72	24,71	-66,15	-10,38	-4,20	29,91	94,79	79,34	84,52	53,72
Alvo 2,0%	107,45	34,01	-11,90	-1,06	-4,89	40,54	173,12	90,00	134,59	90,15
Perda Total (%)	-334,27	-337,33	-603,34	-357,23	-300,69	-307,77	-307,07	-335,47	-321,43	-384,62
Ganho Total (%)	414,90	379,18	598,77	367,48	291,86	335,29	433,13	417,56	402,17	467,07
Retorno Médio (%)	0,15	0,08	-0,01	0,02	-0,02	0,05	0,26	0,17	0,17	0,16
Perda Média (%)	-1,35	-1,29	-2,01	-1,48	-1,13	-1,17	-1,33	-1,43	-1,40	-1,57
Ganho Médio (%)	1,44	1,46	2,24	1,66	1,37	1,26	1,65	1,65	1,68	1,73
Precisão (%)	**53,73**	**49,81**	**47,09**	**47,84**	**44,56**	**50,28**	**53,24**	**51,95**	**51,17**	**52,43**
Acurácia (%)	**52,35**	**49,35**	**48,45**	**52,25**	**48,55**	**50,25**	**55,06**	**54,65**	**53,35**	**55,26**
F1 Score (%)	54,75	50,68	50,91	48,21	45,32	51,70	53,95	52,76	50,74	54,71
Revocação (%)	55,81	51,59	55,39	48,58	46,10	53,20	54,68	53,60	50,31	57,20
Especificidade (%)	48,65	47,07	41,97	55,35	50,65	47,29	55,41	55,60	56,13	53,51
Qtde Perda	248	262	300	242	265	263	231	234	229	245
Qtde Ganho	288	260	267	222	213	266	263	253	240	270
Qtde Total	536	522	567	464	478	529	494	487	469	515
Qtde Previsão Baixa	463	477	432	535	521	470	505	512	530	484
Qtde Previsão Alta	536	522	567	464	478	529	494	487	469	515
Índice *Sharpe* (%)	1,50	-3,15	-8,97	-6,98	-30,90	-4,21	16,51	3,36	3,35	3,01
Índice *Sortino* (%)	2,64	-5,33	-14,79	-11,96	-53,00	-7,21	27,80	5,83	5,54	5,09
Média (%)	0,15	0,08	-0,01	0,02	-0,02	0,05	0,26	0,17	0,17	0,16
Desvio Padrão (%)	1,75	1,78	2,77	2,07	1,61	1,60	1,91	2,02	2,03	2,15
Desvio Absoluto (%)	1,39	1,37	2,12	1,56	1,24	1,22	1,49	1,54	1,54	1,65
Mediana (%)	0,21	0,00	-0,13	-0,11	-0,11	0,04	0,20	0,11	0,09	0,26
Quantil (%)	0,21	0,00	-0,13	-0,11	-0,11	0,04	0,20	0,11	0,09	0,26
DrawDown (%)	-5,05	-9,26	-8,10	-6,32	-4,63	-5,15	-6,17	-7,33	-7,69	-6,86
Máximo (%)	5,79	6,89	10,45	8,89	7,59	9,40	7,50	8,11	6,89	13,04
Amplitude (%)	10,84	16,15	18,55	15,21	12,22	14,55	13,67	15,44	14,58	19,90
Variância (%)	3,06	3,15	7,68	4,30	2,60	2,57	3,64	4,07	4,13	4,62
Preço Inicial (R$)	34,88	35,94	14,11	37,84	9,45	35,00	7,35	15,92	11,08	14,57
Preço Final (R$)	38,68	43,15	35,00	10,68	13,20	37,49	14,85	14,26	17,35	14,28
Distribuição Classe 0 (%)	48,30	49,50	51,80	54,20	53,70	49,90	51,80	52,70	52,30	52,70
Distribuição Classe 1 (%)	51,70	50,50	48,20	45,80	46,30	50,10	48,20	47,30	47,70	47,30
Qtde Real de Baixa	483	495	518	542	537	499	518	527	523	527
Qtde Real de Alta	517	505	482	458	463	501	482	473	477	473
Volatilidade (%)	8,31	8,33	8,10	8,30	2,49	9,64	3,51	4,15	3,77	4,77
Volatilidade *Dowside* (%)	4,75	4,92	4,91	4,84	1,45	5,63	2,08	2,40	2,28	2,81
Shapiro Wilk	0,35	0,00	0,00	0,00	0,00	0,00	0,22	0,00	0,00	0,00
Kolmogorov Smirnov	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
T. Student	0,41	0,68	0,85	0,39	0,94	0,77	0,04	0,15	0,18	0,11
Anova	0,01	0,46	0,64	0,49	0,65	0,31	0,00	0,00	0,00	0,00
Kruskal Wallis	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Tabela 19 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 3 de 5.

Ativos	ECOR3	VALE3	BRKM5	PETR3	PETR4	CSAN3	SBSP3	HYPE3	FLRY3	QUAL3
Retorno Bruto (%)	15,01	-25,20	218,45	67,50	-4,41	147,97	116,29	106,95	136,91	71,05
Custo Operacional (%)	3,30	-5,54	48,06	14,85	-0,97	32,55	25,58	23,53	30,12	15,63
Retorno Líquido (%)	11,71	-19,66	170,39	52,65	-3,44	115,42	90,70	83,42	106,79	55,42
Buy And Hold (%)	10,84	125,80	172,72	211,60	164,64	60,08	127,08	85,75	37,45	-35,14
Comprar Sempre (%)	-72,90	-42,88	86,35	-94,01	-118,80	90,68	130,11	109,99	40,09	-35,71
Inversa (%)	-87,91	-17,68	-132,10	-161,51	-114,39	-57,29	13,83	3,05	-96,82	-106,77
Aleatória (%)	-1,89	27,28	18,19	-37,09	-75,58	25,80	70,94	22,17	29,45	16,65
Alvo 0,5%	-26,03	-18,43	44,03	18,64	43,49	68,44	46,66	70,20	53,56	-3,52
Alvo 1,0%	-27,33	-11,28	78,30	61,18	4,43	82,85	55,27	125,21	115,16	47,23
Alvo 2,0%	-24,00	-41,15	128,41	56,95	-21,95	144,97	126,21	127,06	149,04	65,18
Perda Total (%)	-468,45	-550,91	-354,20	-476,56	-519,28	-332,23	-329,71	-272,11	-296,51	-473,81
Ganho Total (%)	483,47	525,71	572,65	544,06	514,87	480,21	446,00	379,06	433,42	544,86
Retorno Médio (%)	0,03	-0,05	0,46	0,14	-0,01	0,29	0,23	0,20	0,27	0,14
Perda Média (%)	-1,84	-2,10	-1,50	-2,02	-1,97	-1,42	-1,37	-1,09	-1,20	-1,87
Ganho Médio (%)	2,18	1,96	2,43	2,09	2,06	1,76	1,70	1,36	1,63	2,06
Precisão (%)	46,64	50,57	50,00	52,42	48,73	53,85	52,29	52,74	51,75	51,16
Acurácia (%)	50,15	51,45	53,15	55,76	51,85	51,85	51,45	52,45	54,35	52,35
F1 Score (%)	47,13	52,50	50,21	54,05	50,97	53,16	52,03	54,02	53,85	52,68
Revocação (%)	47,64	54,58	50,43	55,79	53,42	52,50	51,77	55,36	56,12	54,30
Especificidade (%)	52,35	48,43	55,56	55,72	50,47	51,15	51,12	49,49	52,76	50,49
Qtde Perda	254	262	236	236	263	234	240	250	248	253
Qtde Ganho	222	268	236	260	250	273	263	279	266	265
Qtde Total	476	530	472	496	513	507	503	529	514	518
Qtde Previsão Baixa	523	469	527	503	486	492	496	470	485	481
Qtde Previsão Alta	476	530	472	496	513	507	503	529	514	518
Índice *Sharpe* (%)	-16,23	-8,17	12,58	-0,09	-11,55	7,39	5,78	5,61	7,74	0,35
Índice *Sortino* (%)	-27,19	-12,89	23,13	-0,14	-17,00	12,70	9,99	9,30	13,85	0,54
Média (%)	0,03	-0,05	0,46	0,14	-0,01	0,29	0,23	0,20	0,27	0,14
Desvio Padrão (%)	2,57	2,65	2,82	2,82	2,75	2,05	1,99	1,72	1,84	2,60
Desvio Absoluto (%)	2,00	2,03	2,00	2,05	2,02	1,60	1,54	1,23	1,43	1,97
Mediana (%)	-0,12	0,09	0,01	0,15	0,00	0,20	0,16	0,15	0,07	0,07
Quantil (%)	-0,12	0,09	0,01	0,15	0,00	0,20	0,16	0,15	0,07	0,07
DrawDown (%)	-9,19	-9,35	-10,60	-16,46	-17,38	-7,48	-8,26	-14,85	-4,78	-13,71
Máximo (%)	9,02	9,62	14,75	10,43	12,31	8,47	6,33	6,77	7,31	7,99
Amplitude (%)	18,21	18,97	25,35	26,89	29,69	15,95	14,59	21,62	12,09	21,70
Variância (%)	6,61	7,03	7,95	7,95	7,57	4,22	3,95	2,97	3,38	6,74
Preço Inicial (R$)	9,88	22,32	16,05	8,30	8,56	25,20	17,06	15,97	15,16	27,49
Preço Final (R$)	11,17	56,15	49,45	29,46	25,54	43,44	43,18	32,47	22,00	15,85
Distribuição Classe 0 (%)	53,30	50,90	53,10	53,40	53,20	47,90	49,10	49,50	52,60	51,20
Distribuição Classe 1 (%)	46,70	49,10	46,90	46,60	46,80	52,10	50,90	50,50	47,40	48,80
Qtde Real de Baixa	533	509	531	534	532	479	491	495	526	512
Qtde Real de Alta	467	491	469	466	468	521	509	505	474	488
Volatilidade (%)	3,27	11,43	11,95	6,88	6,28	10,80	8,34	6,93	8,88	8,33
Volatilidade Dowside (%)	1,95	7,24	6,50	4,47	4,27	6,29	4,82	4,18	4,97	5,41
Shapiro Wilk	0,00	0,02	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00
Kolmogorov Smirnov	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
T. Student	0,46	0,97	0,01	0,14	0,47	0,07	0,36	0,30	0,03	0,21
Anova	0,55	0,88	0,00	0,01	0,24	0,00	0,01	0,00	0,00	0,03
kruskal Wallis	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Tabela 20 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 4 de 5.

Ativos	RADL3	UGPA3	MULT3	WEGE3	IGTA3	RENT3	BBSE3	CVCB3	CSNA3	GGBR4
Retorno Bruto (%)	**61,86**	**64,37**	**99,89**	**45,20**	**103,00**	**122,09**	**52,16**	**104,76**	**-22,95**	**-115,91**
Custo Operacional (%)	13,61	14,16	21,98	9,94	22,66	26,86	11,47	23,05	-5,05	-25,50
Retorno Líquido (%)	48,25	50,21	77,91	35,26	80,34	95,23	40,68	81,71	-17,90	-90,41
Buy And Hold (%)	**121,19**	**13,28**	**-39,01**	**-31,92**	**60,81**	**-5,73**	**-2,26**	**262,78**	**77,28**	**50,38**
Comprar Sempre (%)	-5,85	-46,08	101,61	-68,46	78,22	120,52	16,82	47,19	-209,30	-207,94
Inversa (%)	-67,71	-110,45	1,72	-113,66	-24,78	-1,57	-35,34	-57,57	-186,35	-92,03
Aleatória (%)	-45,76	-33,96	1,33	-33,21	12,79	53,95	-20,19	-26,71	-125,72	-203,13
Alvo 0,5%	2,89	2,15	46,44	24,88	17,64	37,66	0,64	-61,28	-9,48	30,99
Alvo 1,0%	57,76	22,76	111,54	83,40	49,35	93,30	13,87	34,20	18,37	0,22
Alvo 2,0%	84,26	47,27	132,57	65,59	110,69	128,21	46,68	113,64	-9,11	-47,57
Perda Total (%)	-310,02	-294,70	-266,57	-316,37	-287,33	-393,86	-299,74	-363,23	-713,14	-604,95
Ganho Total (%)	371,88	359,07	366,46	361,57	390,33	515,94	351,90	467,99	690,19	489,04
Retorno Médio (%)	0,12	0,12	0,20	0,09	0,20	0,23	0,11	0,21	-0,05	-0,23
Perda Média (%)	-1,25	-1,22	-1,16	-1,21	-1,18	-1,54	-1,22	-1,64	-2,69	-2,12
Ganho Médio (%)	1,36	1,29	1,39	1,48	1,44	1,81	1,57	1,68	2,97	2,24
Precisão (%)	**52,49**	**53,46**	**53,44**	**48,32**	**52,82**	**52,78**	**47,76**	**55,80**	**46,68**	**43,34**
Acurácia (%)	**54,25**	**53,65**	**52,75**	**51,45**	**52,05**	**51,75**	**50,35**	**54,45**	**53,45**	**48,15**
F1 Score (%)	54,53	54,56	52,80	50,26	53,18	54,18	47,46	55,08	49,95	45,70
Revocação (%)	56,73	55,71	52,17	52,35	53,54	55,66	47,16	54,39	53,70	48,34
Especificidade (%)	51,94	51,60	53,35	50,66	50,51	47,64	53,24	54,53	53,26	47,99
Qtde Perda	248	242	230	262	243	255	245	221	265	285
Qtde Ganho	274	278	264	245	272	285	224	279	232	218
Qtde Total	522	520	494	507	515	540	469	500	497	503
Qtde Previsão Baixa	477	479	505	492	484	459	530	499	502	496
Qtde Previsão Alta	522	520	494	507	515	540	469	500	497	503
Índice *Sharpe* (%)	-0,37	-0,25	2,30	-4,49	4,33	4,91	-1,91	3,29	-19,27	-39,26
Índice *Sortino* (%)	-0,62	-0,39	4,11	-7,43	7,63	8,66	-3,31	5,25	-30,98	-63,57
Média (%)	0,12	0,12	0,20	0,09	0,20	0,23	0,11	0,21	-0,05	-0,23
Desvio Padrão (%)	1,69	1,69	1,61	1,72	1,73	2,14	1,78	2,12	3,74	2,81
Desvio Absoluto (%)	1,30	1,25	1,27	1,34	1,31	1,68	1,40	1,65	2,82	2,16
Mediana (%)	0,08	0,11	0,17	-0,06	0,14	0,18	-0,04	0,25	-0,18	-0,27
Quantil (%)	0,08	0,11	0,17	-0,06	0,14	0,18	-0,04	0,25	-0,18	-0,27
DrawDown (%)	-6,65	-6,29	-5,56	-5,41	-6,92	-5,96	-5,59	-7,03	-12,59	-7,15
Máximo (%)	6,16	6,13	5,05	6,95	5,66	8,02	6,98	6,55	15,22	13,32
Amplitude (%)	12,81	12,42	10,61	12,36	12,58	13,98	12,57	13,58	27,81	20,47
Variância (%)	2,86	2,85	2,60	2,94	3,01	4,56	3,16	4,50	14,01	7,89
Preço Inicial (R$)	25,32	49,75	47,36	30,58	24,13	34,61	30,83	14,88	5,37	9,95
Preço Final (R$)	62,29	57,71	25,10	18,82	41,81	32,22	29,99	61,99	10,37	15,99
Distribuição Classe 0 (%)	51,60	50,00	49,30	53,10	49,10	48,80	52,40	48,60	56,80	54,90
Distribuição Classe 1 (%)	48,40	50,00	50,70	46,90	50,90	51,20	47,60	51,40	43,20	45,10
Qtde Real de Baixa	516	500	493	531	491	488	524	486	568	549
Qtde Real de Alta	484	500	507	469	509	512	476	514	432	451
Volatilidade (%)	16,79	15,19	13,84	5,11	8,06	10,98	8,36	11,15	4,72	4,69
Volatilidade Dowside (%)	10,09	9,49	7,73	3,08	4,57	6,23	4,82	6,97	2,94	2,90
Shapiro Wilk	0,01	0,00	0,49	0,03	0,00	0,12	0,00	0,13	0,00	0,00
Kolmogorov Smirnov	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
T. Student	0,19	0,06	0,26	0,10	0,19	0,36	0,35	0,16	0,42	0,89
Anova	0,00	0,00	0,00	0,00	0,00	0,01	0,10	0,00	0,19	0,04
Kruskal Wallis	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Tabela 21 – Modelo *RBM + NB*: Extrato de Validação (1.000 dias de 07/01/2015 a 24/01/2019) - Parte 5 de 5.

Ativos	GOAU4	USIM5	TOTS3	TIMP3	VIVT4	BTOW3	LAME4	LREN3	PCAR4	HGTX3
Retorno Bruto (%)	**-14,53**	**16,10**	**50,99**	**113,66**	**42,02**	**171,83**	**75,00**	**140,55**	**2,25**	**76,32**
Custo Operacional (%)	-3,20	3,54	11,22	25,01	9,24	37,80	16,50	30,92	0,50	16,79
Retorno Líquido (%)	-11,33	12,55	39,77	88,66	32,78	134,03	58,50	109,63	1,76	59,53
Buy And Hold (%)	**-28,34**	**84,71**	**2,00**	**8,21**	**3,23**	**93,98**	**17,19**	**-33,74**	**-1,05**	**47,02**
Comprar Sempre (%)	-300,72	-324,73	-19,46	14,00	-16,35	36,64	-42,51	123,29	-24,32	-34,16
Inversa (%)	-286,19	-340,83	-70,45	-99,66	-58,37	-135,19	-117,51	-17,27	-26,57	-110,48
Aleatória (%)	-238,10	-141,31	-29,81	-1,51	-51,64	9,36	-57,16	65,79	6,78	-37,71
Alvo 0,5%	-22,27	50,92	14,91	20,74	49,83	-36,05	-10,32	60,00	-39,77	-8,96
Alvo 1,0%	-26,47	12,81	68,20	68,43	71,84	27,76	9,56	105,63	-33,48	34,21
Alvo 2,0%	-47,59	-97,32	77,56	108,95	63,08	12,35	43,67	140,99	-22,50	69,36
Perda Total (%)	-576,78	-618,75	-344,93	-330,50	-276,31	-632,50	-367,34	-319,50	-399,51	-424,75
Ganho Total (%)	562,24	634,85	395,91	444,16	318,33	804,33	442,35	460,06	401,76	501,07
Retorno Médio (%)	-0,03	0,04	0,10	0,22	0,08	0,31	0,14	0,28	0,00	0,15
Perda Média (%)	-2,28	-2,55	-1,35	-1,37	-1,14	-2,38	-1,45	-1,38	-1,51	-1,65
Ganho Médio (%)	2,56	3,27	1,62	1,67	1,16	2,77	1,59	1,65	1,52	1,94
Precisão (%)	**46,51**	**44,39**	**48,90**	**52,47**	**53,09**	**52,16**	**52,26**	**54,71**	**50,00**	**50,10**
Acurácia (%)	**52,95**	**53,55**	**51,85**	**53,65**	**53,75**	**53,85**	**53,45**	**53,75**	**51,35**	**53,65**
F1 Score (%)	48,35	45,54	50,36	53,47	54,35	55,72	54,46	54,71	52,07	52,71
Revocação (%)	50,34	46,75	51,91	54,51	55,67	59,79	56,85	54,71	54,32	55,60
Especificidade (%)	54,98	58,39	51,80	52,84	51,88	48,25	50,20	52,76	48,54	51,96
Qtde Perda	253	243	255	241	243	266	254	231	264	257
Qtde Ganho	220	194	244	266	275	290	278	279	264	258
Qtde Total	473	437	499	507	518	556	532	510	528	515
Qtde Previsão Baixa	526	562	500	492	481	443	467	489	471	484
Qtde Previsão Alta	473	437	499	507	518	556	532	510	528	515
Índice *Sharpe* (%)	-30,14	-15,89	-1,72	14,07	-2,47	9,46	1,25	5,43	-3,24	1,17
Índice *Sortino* (%)	-46,39	-25,16	-2,91	22,56	-4,20	16,29	2,02	9,77	-5,49	2,03
Média (%)	-0,03	0,04	0,10	0,22	0,08	0,31	0,14	0,28	0,00	0,15
Desvio Padrão (%)	3,16	4,08	1,95	1,99	1,57	3,41	1,95	1,89	1,98	2,37
Desvio Absoluto (%)	2,41	2,87	1,49	1,52	1,15	2,58	1,52	1,52	1,52	1,80
Mediana (%)	-0,13	-0,43	0,00	0,16	0,08	0,18	0,12	0,23	0,01	0,05
Quantil (%)	-0,13	-0,43	0,00	0,16	0,08	0,18	0,12	0,23	0,01	0,05
DrawDown (%)	-10,30	-13,42	-6,27	-6,84	-4,85	-13,08	-7,58	-7,07	-6,50	-5,63
Máximo (%)	13,39	33,68	7,91	7,93	12,21	13,16	7,36	5,57	6,74	14,69
Amplitude (%)	23,69	47,10	14,18	14,77	17,06	26,24	14,94	12,64	13,24	20,32
Variância (%)	9,99	16,64	3,81	3,97	2,46	11,64	3,82	3,58	3,92	5,64
Preço Inicial (R$)	11,45	4,85	32,81	11,42	45,98	21,39	16,95	74,88	96,00	18,80
Preço Final (R$)	7,54	9,80	33,60	12,55	47,77	45,61	20,46	44,44	94,78	29,45
Distribuição Classe 0 (%)	56,30	58,40	53,00	51,20	50,50	51,40	51,10	48,90	51,30	53,50
Distribuição Classe 1 (%)	43,70	41,60	47,00	48,80	49,50	48,60	48,90	51,10	48,70	46,50
Qtde Real de Baixa	563	584	530	512	505	514	511	489	513	535
Qtde Real de Alta	437	416	470	488	495	486	489	511	487	465
Volatilidade (%)	2,74	3,27	9,95	3,24	10,55	10,96	5,50	13,34	20,31	7,01
Volatilidade Dowside (%)	1,78	2,07	5,89	2,02	6,21	6,37	3,42	7,41	11,99	4,03
Shapiro Wilk	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,74	0,01	0,00
Kolmogorov Smirnov	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
T. Student	0,14	0,11	0,27	0,06	0,24	0,14	0,09	0,14	0,79	0,16
Anova	0,01	0,00	0,01	0,00	0,00	0,01	0,01	0,00	0,86	0,01
Kruskal Wallis	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00