

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM DE SISTEMAS COMPLEXOS

SERGIO REINALDO MARTELETTO

Técnicas de seleção de atributos através de *Random Forests*:
um estudo de caso para detecção de tendências em séries temporais financeiras

São Paulo

2022

SERGIO REINALDO MARTELETTO

Técnicas de seleção de atributos através de *Random Forests*:

um estudo de caso para detecção de tendências em séries temporais financeiras

Versão Corrigida

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo como parte dos requisitos para obtenção do título de Mestre em Ciências pelo programa de pós-graduação em Modelagem de Sistemas Complexos.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Marcelo de Souza Laretto

São Paulo

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha catalográfica elaborada pela Biblioteca da Escola de Artes, Ciências e Humanidades,
com os dados inseridos pelo(a) autor(a)
Brenda Fontes Malheiros de Castro CRB 8-7012; Sandra Tokarevicz CRB 8-4936

Marteletto, Sergio Reinaldo
Técnicas de seleção de atributos através de
Random Forests: um estudo de caso para detecção de
tendências em séries temporais financeiras / Sergio
Reinaldo Marteletto; orientador, Marcelo de Souza
Lauretto. -- São Paulo, 2022.
85 p: il.

Dissertacao (Mestrado em Ciencias) - Programa de
Pós-Graduação em Modelagem de Sistemas Complexos,
Escola de Artes, Ciências e Humanidades,
Universidade de São Paulo, 2022.
Versão corrigida

1. aprendizado de máquina. 2. random forests. 3.
seleção de atributos. 4. séries temporais. I.
Lauretto, Marcelo de Souza, orient. II. Título.

Avaliação

Nome: Sergio Reinaldo Marteletto

Título: **Técnicas de Seleção de Atributos através de Random Forests:**
um estudo de caso para detecção de tendências em séries temporais
financeiras.

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da
Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Aprovado em 11/05/2022

Banca Examinadora

Prof. Dr.	Fernando Fagundes Ferreira
Instituição	Universidade de São Paulo

Prof. Dr.	Norton Trevisan Roman
Instituição	Universidade de São Paulo

Prof. Dr.	Ricardo Augusto Souza Fernandes
Instituição	Universidade Federal de São Carlos

Dedicatória

ao meu melhor amigo, Marcos Antonio Alves.

Agradecimentos

Agradeço primeiramente à Escola de Artes, Ciências e Humanidades, da Universidade de São Paulo, pela oportunidade de realização do mestrado.

Agradeço especialmente ao meu orientador, Prof. Dr. Marcelo de Souza Lauretto, pela amizade, paciência e ensinamentos, sem os quais esse trabalho de pesquisa não se concretizaria.

Pelo repertório de experiências adquirido, agradeço ao corpo docente que manteve contato durante a convivência no campus.

Pela prestação de serviços, sempre cordial, agradeço aos funcionários e funcionárias da biblioteca, da manutenção, do refeitório e da secretaria de pós-graduação.

Pelo apoio financeiro concedido, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, além da inestimável contribuição de Guilherme Rafare.

Pelo suporte computacional, agradeço ao Centro de Ciências Matemáticas Aplicadas à Indústria – CEPID/CeMEAI (projeto FAPESP 2013/07375-0).

Por fim, eu também quero agradecer à Margô Isaac, uma mulher generosa e especial.

“Essencialmente, todos os modelos estão errados,
mas alguns são úteis”.

George E. P. Box (estatístico britânico)

Resumo

MARTELETTO, Sergio Reinaldo. **Técnicas de seleção de atributos através de *Random Forests***: um estudo de caso para detecção de tendências em séries temporais financeiras. 2022. 85 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2022.

Nas últimas décadas tem havido um interesse crescente em prever o comportamento futuro dos mercados financeiros. Pesquisadores investigam esse problema modelando uma representação conveniente para os dados, as chamadas séries temporais, apesar da dificuldade de estudá-las com precisão devido aos seus padrões não lineares e não estacionários. Além disso, a questão da alta dimensionalidade, presente no conjunto de dados, reduz o entendimento das relações de dependência entre as observações. O uso de novas tecnologias em finanças, como o aprendizado de máquina, busca extrair e analisar informações sobre o preço dos ativos e fluxos de negociação em um ambiente competitivo de risco-retorno. Esse trabalho propõe a análise comparativa de técnicas modernas de seleção de atributos – VSURF (*Variable Selection Using Random Forests*) e RFE (*Recursive Feature Elimination*), a fim de reduzir a dimensionalidade na base de dados. Os resultados obtidos foram consistentes e não causaram perda da capacidade preditiva do modelo.

Palavras-chave: Aprendizado de Máquina. *Random Forests*. Seleção de Atributos. Séries Temporais.

Abstract

MARTELETTO, Sergio Reinaldo. **Feature selection techniques using Random Forests**: a case study for detecting trends in financial time series. 2022. 85 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2022.

In recent decades there has been a growing interest in predicting the future behavior of financial markets. Researchers investigate this problem by modeling a convenient representation of the data, the so-called time series, despite the difficulty of studying them accurately due to their non-linear and non-stationary patterns. In addition, the issue of high dimensionality present in the dataset reduces the understanding of the dependence relationships between the observations. The use of new technologies in finance, such as machine learning, seeks to extract and analyze information about asset prices and trading flows in a competitive risk-return environment. This work proposes a comparative analysis of modern feature selection techniques – VSURF (Variable Selection Using Random Forests) and RFE (Recursive Feature Elimination), in order to reduce the dimensionality present in the dataset. The results obtained were consistent and did not cause predictive ability of the model.

Keywords: Machine Learning. Random Forests. Feature Selection. Time Series.

Lista de Figuras

Figura 1 – Abrangência da ciência da complexidade.....	2
Figura 2 – Tipos de aprendizado de máquina	4
Figura 3 – Componentes de uma série temporal.....	8
Figura 4 – Direções das tendências de mercado.....	14
Figura 5 – Identificação das tendências primária, secundária e terciária.....	15
Figura 6 – Bandas de Bollinger BB.....	18
Figura 7 – Índice de Movimento Direcional ADX.....	19
Figura 8 – Parabólico SAR.....	20
Figura 9 – Médias móveis Convergentes / Divergentes MACD.....	21
Figura 10 – Índice do Canal de Commodities CCI.....	22
Figura 11 – Índice de Força Relativa RSI.....	23
Figura 12 – Oscilador Estocástico STO.....	24
Figura 13 – On-balance Volume OBV.....	25
Figura 14 – Linha de Acumulação / Distribuição A/D.....	26
Figura 15 – Preço médio ponderado pelo volume VWAP.....	27
Figura 16 – Representação de um sumário de livro numa árvore de decisão.....	29
Figura 17 – Conjunto E de imagens de objetos capturados por câmera.....	30
Figura 18 – Árvore de decisão induzida a partir do conjunto E.....	31
Figura 19 – Comparação entre medidas de impureza de um nó.....	35
Figura 20 – Técnica k-Fold de validação cruzada.....	38
Figura 21 – Técnica Holdout de validação cruzada.....	39
Figura 22 – Matriz de confusão.....	40
Figura 23 – Métodos de seleção de atributos.....	43

Figura 24 – Etapas do método de seleção de variáveis VSURF.....	47
Figura 25 – Evolução do Ibovespa, principal índice de ações da B3.....	49
Figura 26 – Evolução de três médias móveis sobre o preço de um ativo.....	53
Figura 27 – Capacidade discriminante por classe de cada atributo comum.....	65

Lista de Tabelas

Tabela 1 – Vetores de atributos dos objetos do conjunto E.....	30
Tabela 2 – Comparação de características entre métodos de seleção.....	44
Tabela 3 – Rotulagem da variável resposta do conjunto de dados.....	54
Tabela 4 – Proporções entre as classes nos conjuntos de dados.....	55
Tabela 5 – Número de atributos selecionados e tempos de processamento.....	58
Tabela 6 – Métricas de desempenho global.....	59
Tabela 7 – Métricas de desempenho por classe.....	60
Tabela 8 – Identificação dos principais atributos selecionados.....	62
Tabela 9 – Estabilidade dos atributos entre os métodos de seleção.....	64

Sumário

1. INTRODUÇÃO	1
1.1 Sistemas Complexos e Inteligência Artificial.....	1
1.2 Aprendizado de Máquina e Mercados Financeiros.....	4
1.3 Séries Temporais e Dimensionalidade nos Dados.....	8
1.4 Objetivos.....	11
1.5 Organização.....	11
2. ANÁLISE TÉCNICA, TENDÊNCIAS e INDICADORES FINANCEIROS	12
2.1 Considerações sobre a análise técnica.....	12
2.2 Tendências de mercado.....	14
2.3 Indicadores financeiros.....	16
2.3.1 Rastreadores.....	18
2.3.2 Osciladores.....	21
2.3.3 Indicadores de Volume.....	25
3. ÁRVORES de DECISÃO, <i>RANDOM FORESTS</i> e DESEMPENHO	28
3.1 Árvores de Decisão	28
3.1.1 Construção de uma árvore de decisão.....	29
3.1.2 Critérios para o particionamento dos nós.....	32
3.2 <i>Random Forests</i>	36
3.3 Desempenho dos Classificadores.....	37
3.3.1 Validação Cruzada.....	37
3.3.2 Matriz de Confusão.....	39
3.3.3 Métricas de Desempenho	41

4. MÉTODOS DE SELEÇÃO DE ATRIBUTOS	42
4.1 Importância da Seleção de Atributos.....	42
4.2 VSURF – <i>Variable Selection Using Random Forests</i>	45
4.3 RFE – <i>Recursive Feature Elimination</i>	47
5. METODOLOGIA	49
5.1 Caracterização do conjunto de dados inicial.....	49
5.2 Atributos base – a primeira engenharia de atributos.....	50
5.3 Atributos expandidos – a segunda engenharia de atributos.....	51
5.4 Rotulagem das classes na variável resposta.....	52
5.5 Divisão do conjunto de dados e parametrização dos métodos.....	55
6. RESULTADOS	57
6.1 Comparação dos métodos e análise das métricas de desempenho.....	57
6.2 Estabilidade e capacidade discriminante dos atributos selecionados.....	62
7. CONCLUSÕES	67
REFERÊNCIAS	69

Capítulo 1

Introdução

Neste capítulo, apresentamos o contexto geral no qual o trabalho está inserido. Inicialmente, introduzimos conceitos sobre a ciência da complexidade e sua relação com a inteligência artificial. Abordamos também os tipos de aprendizado de máquina e sua aplicação na análise dos mercados financeiros. Por fim, descrevemos algumas definições sobre séries temporais e a questão da alta dimensionalidade nos dados.

1.1 Sistemas Complexos e Inteligência Artificial

Nas últimas décadas foi possível notar que o crescimento urbano, o movimento das economias e o avanço de novas tecnologias nos conduziu à ideia de um mundo cada vez mais complexo, uma vez que a velocidade das mudanças é maior que a capacidade de compreendê-las. Neste contexto, a teoria da complexidade surgiu para auxiliar na compreensão dos princípios e mecanismos que regem os chamados sistemas complexos, como o tráfego nas grandes cidades, a proliferação de doenças contagiosas, os mercados financeiros, entre outros (BAR-YAM, 2003).

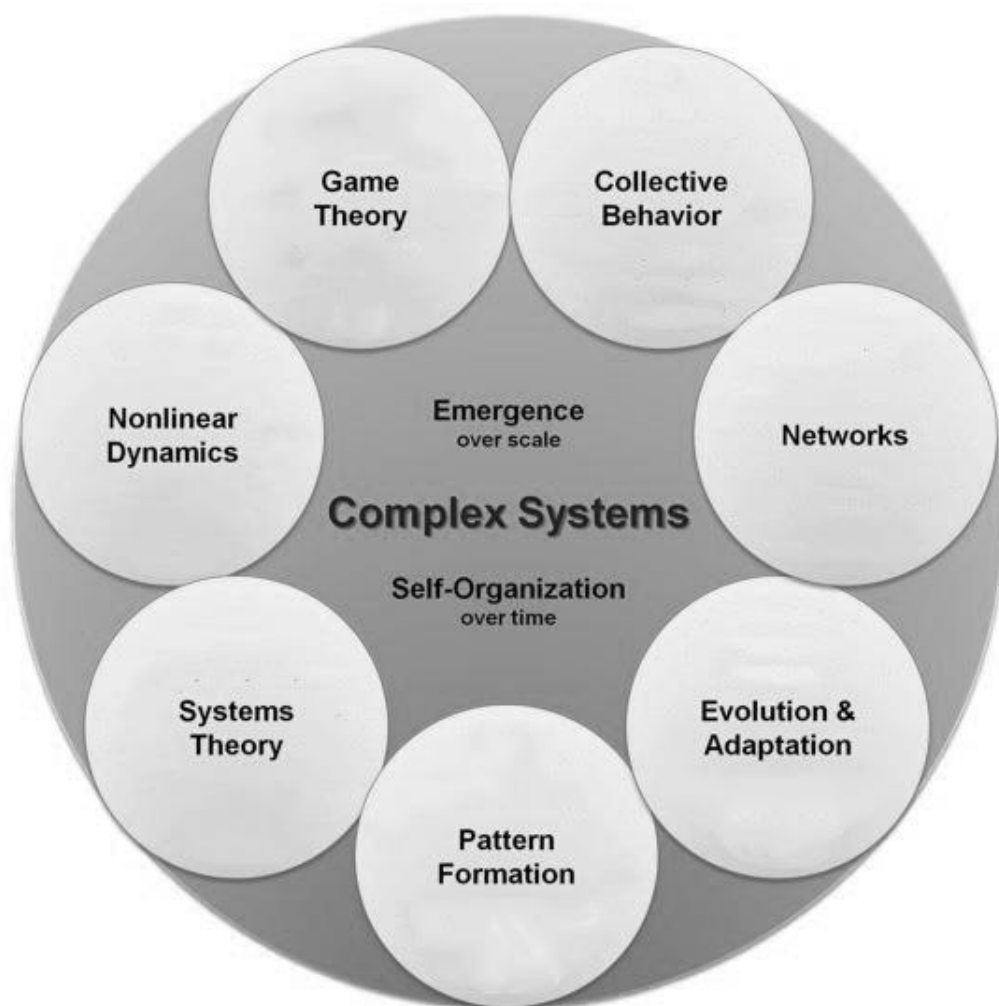
Esses fenômenos que nos envolvem são ditos “complexos” por representarem sistemas físicos que desafiam o emprego de modelos na previsão do comportamento. Os sistemas complexos são caracterizados pelo surgimento da emergência, da auto-organização e da adaptabilidade a modificações externas. Em tais sistemas, seus componentes interagem entre si e produzem manifestações inesperadas e não lineares, podendo surgir propriedades que não estavam presentes nos elementos individuais. Tal fenômeno de aparecimento de propriedades de comportamento coletivo é denominado emergência (TAYLOR, 2003).

Além da emergência, um outro conceito fundamental no estudo dos sistemas complexos é o de auto-organização. A “organização” refere-se ao processo que gera um aumento na ordem ou na estrutura dos componentes do sistema, podendo promover uma função específica, sem a presença de um controle central ou externo. Um sistema auto-organizado pode sofrer influências do meio ao seu redor, no sentido

de introduzir ruídos ou perturbações, mas é capaz de se adaptar para continuar exercendo sua função. Essa capacidade de adaptação é conhecida como robustez, uma propriedade importante dos sistemas auto-organizados (TAYLOR, 2003).

Em torno desses conceitos-chave de emergência e auto-organização, foi criado um mapa visual que propõe a divisão da ciência dos sistemas complexos em sete áreas principais, conforme a Figura 1. Vale ressaltar que as raízes históricas da teoria da complexidade estão representadas pelos três grupos à esquerda: Teoria dos Jogos, Dinâmica não Linear e Teoria de Sistemas. Já os demais grupos representam os campos de estudos mais recentes: Comportamento Coletivo, Redes, Evolução e Adaptação, além da Formação de Padrões (SAYAMA, 2015).

Figura 1 – Abrangência da ciência da complexidade



Fonte: adaptado de Sayama (2015)

Especificamente, a área de Evolução e Adaptação foi inicialmente discutida no âmbito da biologia evolutiva, que pode ser rastreada até a teoria de Charles Darwin. Mas outro contexto, que é mais próximo à teoria da complexidade, é o de sistemas adaptativos complexos e que envolve computação evolutiva, redes neurais artificiais e outras estruturas inspiradas em processos biológicos e neurológicos. Este grupo da Evolução e Adaptação abrange também a inteligência artificial e o aprendizado de máquina, que são os campos de pesquisa que motivam este trabalho.

Até o final do século XX, a inteligência artificial (IA) era vista como um campo teórico, com aplicações em alguns problemas desafiadores, mas de pouco valor prático. A partir da década de 1970, houve uma disseminação do uso de técnicas de computação baseadas em IA para a solução de problemas reais. Esses problemas eram tratados, computacionalmente, por meio da aquisição de conhecimento de especialistas num dado domínio. Tais programas de computador eram conhecidos como sistemas especialistas ou baseados em conhecimento (FACELI *et al.*, 2011).

Neste processo de aquisição de conhecimento eram realizadas entrevistas com os especialistas para descobrir quais regras eram utilizadas para tomada de decisão. Tal processo possuía várias limitações como, por exemplo, a subjetividade decorrente da intuição de cada profissional no momento da decisão, além da pouca cooperação individual, em virtude do receio de ser dispensado após repassar o conhecimento solicitado (FACELI *et al.*, 2011).

O livro *Probabilistic Reasoning in Intelligent System*, publicado em 1988, levou a uma nova aceitação da probabilidade e da teoria da decisão em IA. O formalismo denominado rede bayesiana foi criado para permitir a representação eficiente do conhecimento incerto e do raciocínio rigoroso, a partir da utilização desse tipo de conhecimento. Essa abordagem supera vários problemas dos sistemas de décadas anteriores, já que admite o aprendizado a partir da experiência e combina o melhor da IA clássica com novas técnicas computacionais (PEARL, 1988).

Por exemplo, grande parte do trabalho com redes neurais artificiais, a partir da década de 1990, foi realizada na tentativa de compreender como essas redes se diferenciavam das técnicas tradicionais. Atualmente, esse campo chegou a tal compreensão que, agora, as redes neurais abrangem técnicas da estatística, do reconhecimento de padrões e do aprendizado de máquina. Como será visto na

próxima seção, o resultado desse desenvolvimento, em conjunto com a expansão do poder computacional e a geração massiva de dados, culminou no surgimento de uma nova e vigorosa indústria (RUSSELL; NORVIG, 2016).

1.2 Aprendizado de Máquina e Mercados Financeiros

Com a crescente complexidade dos problemas ficou evidente a necessidade de ferramentas computacionais mais sofisticadas, e que fossem mais autônomas, de forma a reduzir a necessidade da intervenção humana. Essas técnicas deveriam ser capazes de criar por si próprias, a partir de experiência passada, uma hipótese (ou aproximação de função) para resolver um determinado tipo de problema. A este processo de indução de hipótese, dá-se o nome de aprendizado de máquina (AM) (MITCHELL, 2017).

Com os algoritmos de AM, a inteligência artificial vem se desenvolvendo, através de novas técnicas de programação, e sendo capaz de se aprimorar por meio de treinamento, ou seja, seus algoritmos processam grandes quantidades de dados e, dessa forma, podem aprender com os dados fornecidos (SARKER, 2021). Basicamente, é possível classificar o aprendizado de máquina em quatro tipos: aprendizado supervisionado, não supervisionado, semi supervisionado e aprendizado por reforço, conforme apresentado na Figura 2.

Figura 2 – Tipos de Aprendizado de Máquina



No aprendizado supervisionado, que é o foco desse trabalho, o algoritmo é construído a partir de um conjunto de dados que fornece tanto os atributos de entrada (variáveis preditoras) quanto o atributo de saída (variável resposta), ou seja, o modelo é treinado com todas as observações devidamente rotuladas. A partir daí, o algoritmo busca associações entre os parâmetros fornecidos, para que ao final das iterações encontre uma função matemática que melhor relacione as variáveis de entrada e saída (TAEHO, 2021).

O aprendizado supervisionado pode ser separado em duas tarefas principais:

- Classificação: os modelos fornecem uma saída discreta, ou seja, os dados de teste são rotulados numa classe, ou categoria. Por exemplo: tendências do mercado financeiro: 1 (alta), -1 (baixa) ou 0 (lateral);
- Regressão: os modelos fornecem uma saída contínua, ou seja, os dados de teste são rotulados com um valor numérico. Por exemplo: preço estimado de uma determinada ação: 85,90.

No aprendizado não supervisionado, o algoritmo não recebe os atributos de saída, ou seja, no conjunto de dados fornecido ao modelo a variável resposta não está rotulada, de modo que as relações entre as observações são percebidas pelo algoritmo de forma abstrata, sem a necessidade de intervenção humana. Essa característica do aprendizado não supervisionado confere bastante versatilidade ao modelo, uma vez que podem se adaptar aos dados alterando de forma dinâmica as estruturas ocultas (TAEHO, 2021).

O aprendizado semi supervisionado trabalha com conjuntos de dados que possuem atributos de saída rotulados e não rotulados. Como dito anteriormente, dados rotulados são aqueles cuja variável de saída possui uma classe ou valor já definido, ou seja, uma informação significativa para que o algoritmo possa entender a relação entre os parâmetros de entrada e saída. Dados não rotulados não possuem essa informação. Ao usar essa combinação, os modelos de aprendizado semi supervisionado podem aprender a classificar aqueles dados que não foram rotulados (TAEHO, 2021).

O aprendizado por reforço se inspira diretamente no conceito psicológico de condicionamento, onde o algoritmo se aprimora e aprende utilizando um método de tentativa e erro. Nesse tipo de aprendizado o modelo é construído num ambiente tendo

um interpretador e um sistema de recompensa. A cada iteração, o interpretador avalia a saída e, se o resultado for favorável, fornece uma recompensa ao algoritmo. Caso contrário, o modelo é punido e forçado a reiterar até encontrar um resultado melhor. Um caso típico do aprendizado por reforço é a personalização de ofertas on-line, onde o algoritmo reconhece o perfil de cada cliente e sugere preços que terão mais chance de propiciar uma venda bem-sucedida (TAEHO, 2021).

Embora o AM esteja naturalmente associado à IA, outros campos de pesquisa são importantes e têm contribuições diretas e significativas para o avanço do AM, tais como: probabilidade e estatística, econometria, teoria da informação, entre outros. O AM é uma das áreas computacionais que mais tem crescido nos últimos anos. Existem várias aplicações para seus algoritmos, como a condução autônoma de automóveis, a detecção de fraude no uso de cartões de crédito e a análise do comportamento de diferentes ativos, de ações a *commodities* (MITCHELL, 2019).

Os mercados financeiros desempenham um papel importante no crescimento econômico de um país. Se ao longo de um ano o valor do índice das principais ações listadas na bolsa de valores fechar em alta, é bem provável que tal economia venha a crescer. Mercados e ações individuais se movem em tendências e ciclos econômicos tendem a se repetir. A correta previsão dessas tendências, seja de curto ou longo prazo, é de grande importância para os investidores, já que facilita o gerenciamento de risco e retorno das operações financeiras (KIRKPATRICK II; DAHLQUIST, 2010).

As mais recentes tecnologias proporcionam aos investidores acesso a diversos mercados e diferentes ativos, para que possam executar suas decisões de compra e venda, além de utilizar ferramentas de análise de preços e/ou volume. A aplicação do aprendizado de máquina permite explorar novas oportunidades de investimentos em renda variável. Entre os principais métodos de aprendizagem supervisionada, as *Random Forests* vêm se destacando na previsão de tendências em diferentes mercados acionários (MANOJLOVIC; STAJDUHAR, 2015).

A forma de investir também mudou pelo menos sob duas perspectivas. Primeiro, a proliferação de modernas ferramentas e a velocidade de acesso aos dados em tempo real permitem agora a rápida tomada de decisão, além da execução através de softwares de negociação multiplataforma. Em segundo lugar, a IA apoiada pelos avanços tecnológicos, levou a uma nova forma de investir, onde as decisões podem

ser tomadas através do tratamento de grandes quantidades de informação e a quantificação dos riscos envolvidos pode ser feita com muito mais precisão (ARÉVALO et al., 2017).

O crescente acesso global a grandes bases de dados com preços e volumes, e os indicadores técnicos obtidos a partir deles, permite que as decisões sejam tomadas reduzindo o componente emocional, e apostando num enquadramento mais racional baseado na análise da informação quantitativa. Outras fontes de informações, como o sentimento de mercado, também podem ser adicionadas aos preços. Dessa forma, as técnicas de aprendizado de máquina exercem uma contribuição destacada, facilitando a decisão racional e viabilizando oportunidades lucrativas de investimentos (NGUYEN; SHIRAI; VELCIN, 2015).

Existe uma vasta literatura onde os pesquisadores usam preços históricos para prever o comportamento das ações, e também para tentar descobrir padrões nos dados por meio do aprendizado de máquina. Os algoritmos de *Random Forests* vêm sendo amplamente utilizados para seleção de ações (ZHENG et al., 2019), bem como para previsão em séries temporais (TYRALIS; PAPACHARALAMPOUS, 2017).

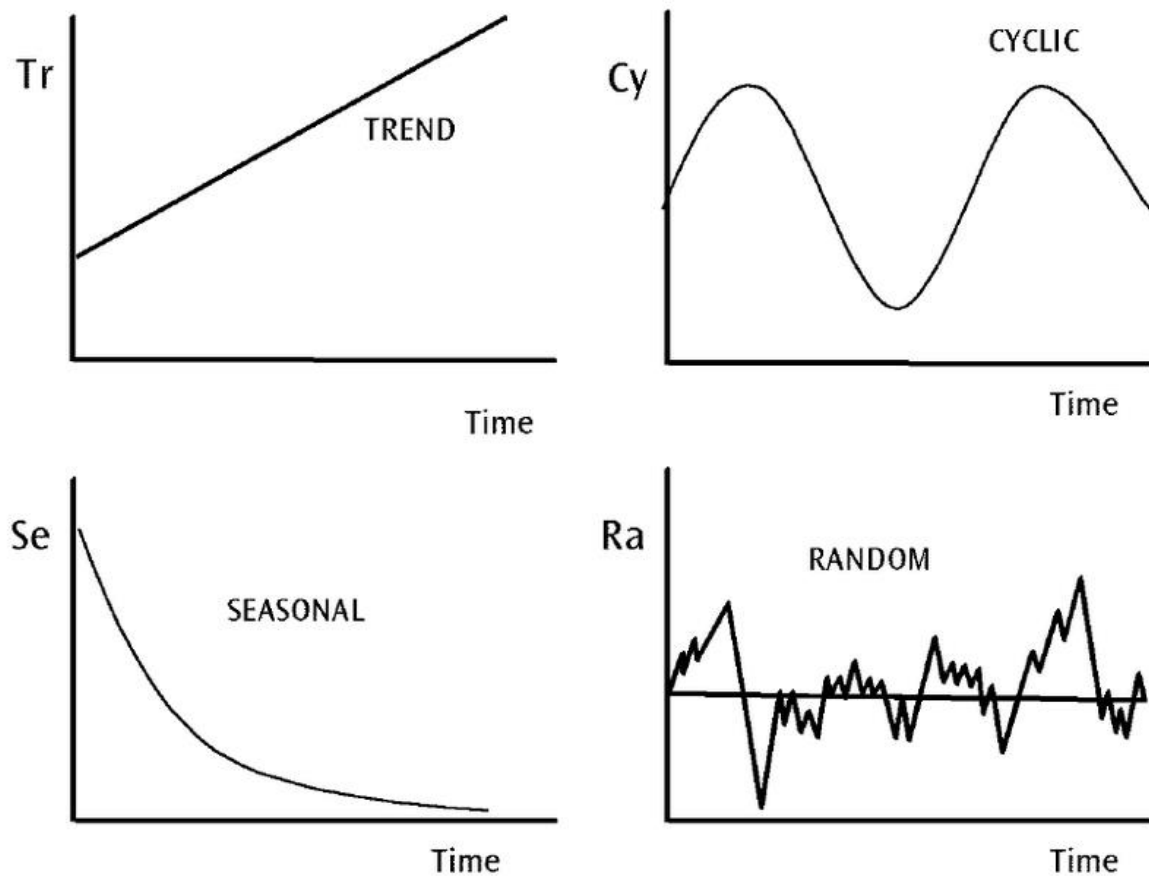
Atualmente, a maioria das plataformas de negociação permite acompanhar ações, índices futuros, moedas e *commodities*. Tais plataformas também incluem indicadores técnicos – rastreadores e osciladores, que são cálculos estatísticos obtidos a partir desses preços históricos e dos volumes negociados, e usados na previsão de tendências (PATEL et al., 2015).

A quantidade de indicadores é enorme e, em muitos casos, os sinais oferecidos por alguns se contradizem com os sinais de outros. Essa análise de rastreadores e osciladores visa a dar suporte na identificação das melhores oportunidades de entrada ou saída do mercado, exigindo que os investidores gerenciem muita informação em tempo real. Nesse contexto, os métodos de aprendizado de máquina podem fornecer soluções para as quais os investidores não conseguiriam lidar com eficiência, como analisar simultaneamente diferentes indicadores, e suas interações, de modo a identificar mudanças sutis no comportamento do mercado (WENG et al., 2020).

1.3 Séries Temporais e Dimensionalidade nos Dados

Uma série temporal é uma coleção de pontos de dados coletados em intervalos de tempo, que são analisados para determinar a tendência de longo prazo, a fim de se realizar uma previsão. O tempo não é uma variável dos dados, mas sim o índice dos dados (TSAY, 2002). Uma série temporal pode conter os seguintes componentes, conforme apresentado na Figura 3.

Figura 3 – Componentes de uma série temporal



Fonte: Analytics Vidhya (2021)

a) Componente Tendência: a tendência é o componente de longo prazo de uma série temporal. Uma tendência pode ser positiva ou negativa, dependendo se a série exibe um padrão crescente ou decrescente no longo prazo. Se não ocorrer um padrão, a série temporal é considerada estacionária;

b) Componente Sazonalidade: flutuações sazonais descrevem qualquer variação regular durante um determinado mês (ou meses), ao longo do ano. Tais variações são causadas pelo clima, costumes sociais, tipo de negócio, entre outros;

c) Componente Ciclo: qualquer padrão mostrando um movimento para cima e para baixo em torno de uma determinada tendência é identificado como um padrão cíclico. A variação cíclica é um componente não sazonal e apresenta natureza periódica com quatro fases: pico, recessão, depressão e expansão;

d) Componente Irregular ou Ruído: são variações imprevisíveis que ocorrem devido a causas repentinas, também chamadas de flutuações acidentais ou erráticas. Toda série temporal apresenta um componente imprevisível e, quando as variações cíclicas e tendências são removidas de um conjunto de dados, o restante residual pode ou não ser aleatório.

Nem todas as séries temporais são iguais, pois existem diferentes fenômenos que podem ser observados. Elas podem ser usadas na previsão do clima, de vendas, do tráfego, de demanda, entre outros. As séries temporais podem ser classificadas como: univariada x multivariada; estacionária x não estacionária; linear x não linear; e sazonal x não sazonal (TSAY, 2002).

Os preços de ativos financeiros produzem uma série temporal financeira. Atualmente, todas as transações de um ativo são registradas – abertura, fechamento, máxima, mínima e volume negociado, levando a uma grande quantidade de dados disponíveis. As incertezas estocásticas inerentes às séries temporais financeiras tornam o assunto bastante interessante para economistas e matemáticos, pois permite fazer inferências e previsões.

A análise de séries temporais financeiras envolve coletar dados históricos, analisar padrões e relacionamentos entre eles e produzir resultados que facilitem a previsão de eventos futuros aplicados em situações semelhantes. Por exemplo, considerando os preços de um índice de ações, é possível modelar suas informações históricas, de forma a avaliar o comportamento atual desse índice. Isso implica que, eventualmente, poderá existir alguma relação entre os preços negociados hoje e os preços futuros (TAYLOR, 1986).

A modelagem de séries temporais tem sido uma área-chave da pesquisa acadêmica, sendo parte integrante de aplicações em ciências biológicas, mudanças climáticas, além da tomada de decisões em finanças. Os métodos modernos de modelagem, através do aprendizado de máquina, fornecem um sistema dinâmico temporal e puramente orientado a dados. Com o crescente poder computacional e a maior disponibilidade de dados, o aprendizado de máquina tornou-se uma parte vital dos modelos de previsão de séries temporais (AHMED, 2010).

A questão da dimensionalidade refere-se aos problemas que surgem quando o conjunto de dados apresenta alta dimensão. A dimensão de um conjunto corresponde ao número de atributos (ou variáveis) que existe para cada observação (ou instância). Um conjunto de dados com um grande número de atributos, da ordem de cem variáveis ou mais, é considerado de alta dimensão. Nesses casos, os tipos mais comuns de problemas que podem se manifestar são: a dificuldade de identificar padrões durante a análise dos dados e/ou de realizar o treinamento de modelos de aprendizado de máquina. Tais problemas também são referidos como a “maldição da dimensionalidade” (FAN; LI, 2006).

Outro ponto importante a se destacar, diz respeito ao *overfitting*, ou seja, quando o modelo fica super ajustado ao conjunto de treinamento, apresentando baixo desempenho ao ser aplicado no conjunto de teste ou com dados reais. Evitar o *overfitting* é uma das principais motivações para a redução da dimensionalidade, uma vez que uma menor quantidade de variáveis no conjunto de treinamento significa menos suposições feitas pelo modelo (DONOHO, 2000). A redução de atributos também apresenta outras vantagens: remove variáveis redundantes, consome menor espaço de armazenamento e implica em menor custo computacional com algoritmos mais rápidos (YAOHAO, 2019).

O processo de redução de dimensionalidade pode ser feito tanto por técnicas de seleção de atributos quanto por métodos de engenharia de variáveis. A seleção de atributos trata da identificação e extração das características mais relevantes para o conjunto de dados. A engenharia de variáveis é a criação de novas características, a partir de atributos já existentes, através de alguma transformação específica. Entre os métodos lineares mais tradicionais de redução de dimensionalidade, é possível citar: a análise fatorial, a análise de componentes principais e a análise discriminante linear (CUNNINGHAM; GHAHRAMANI, 2015). Com relação aos métodos não lineares, se

destacam: a escala multidimensional, o mapeamento de recursos isométricos e a incorporação espectral (LEE; VERLEYSEN, 2021). Diante disso, analisar novas técnicas de seleção de atributos se torna importante para lidar com a “maldição da dimensionalidade”.

1.4 Objetivos

O objetivo geral deste trabalho é realizar uma avaliação empírica comparativa entre os métodos de seleção VSURF (*Variable Selection Using Random Forests*) e RFE (*Recursive Feature Elimination*), para reduzir a dimensionalidade de um conjunto de dados, dentro do contexto de previsão de tendências do Ibovespa, que é o principal índice de ações da bolsa de valores brasileira (B3).

Os objetivos específicos são analisar o impacto desses métodos de seleção sobre a capacidade preditiva das *Random Forests*, e também verificar a persistência dos atributos que foram selecionados em cada um dos métodos.

1.5 Organização

No capítulo 2 são apresentadas algumas considerações sobre análise técnica, sobre como identificar as tendências de mercado, além das características dos indicadores financeiros utilizados neste trabalho. O capítulo 3 descreve as árvores de decisão, os algoritmos de *Random Forests* e os parâmetros de desempenho dos classificadores. O capítulo 4 apresenta os métodos de seleção de variáveis – VSURF e RFE, que foram utilizados. No capítulo 5 é apresentada a metodologia, desde o tratamento inicial do conjunto de dados até a modelagem preditiva com as *Random Forests*. O capítulo 6 apresenta e discute os resultados obtidos pelos métodos de seleção, as métricas de desempenho, o ranking dos atributos selecionados e uma análise do grau de estabilidade e da capacidade discriminante. Finalmente, o capítulo 7 apresenta as conclusões deste trabalho e sugere possíveis desdobramentos futuros.

Capítulo 2

Análise Técnica, Tendências e Indicadores Financeiros

Neste capítulo são apresentadas considerações importantes sobre a análise técnica e sobre as tendências de mercado. Além disso, também serão apresentadas as principais características dos indicadores financeiros utilizados neste trabalho.

2.1 Considerações sobre a Análise Técnica

Existem, basicamente, duas abordagens para se analisar o comportamento de um ativo negociado na bolsa de valores: a análise técnica e a fundamentalista, sendo que a principal diferença está nos fatores que cada uma leva em consideração. De forma simplificada, a análise técnica está relacionada com o *timing*, o momento certo de comprar ou vender determinado ativo. Já a análise fundamentalista é recomendada para tomada de posições de longo prazo. Vale ressaltar que não se trata de escolher uma e negar a outra, pois ambas podem coexistir e, em determinados contextos, até serem complementares (MURPHY, 1999).

A análise fundamentalista utiliza fatores como as demonstrações contábeis, a taxa de juros, o risco-país e projeções macro e microeconômicas, para determinar um fluxo de caixa futuro e trazê-lo a valor presente, a fim de projetar o preço-alvo de uma ação. Este tipo de análise cria uma diferença entre “o preço que deveria ser” e “o preço que realmente é”, uma vez que não leva em consideração a expectativa dos grandes investidores (bancos, institucionais e estrangeiros), e que são os responsáveis pelo movimento dos preços (KIRKPATRICK II; DAHLQUIST, 2010).

A análise técnica avalia o comportamento do mercado através do estudo de padrões gráficos, com o objetivo de prever a tendência futura dos preços, tanto de curto quanto de longo prazo. Este comportamento se sustenta, basicamente, sobre três informações que são disponibilizadas pelas bolsas de valores: preço, volume e contratos em aberto (KIRKPATRICK II; DAHLQUIST, 2010). A análise técnica considera ainda as seguintes premissas:

- a) o comportamento do mercado desconta tudo – as variáveis fundamentais, políticas, psicológicas ou de qualquer outra ordem, já estão refletidas no preço dos ativos;
- b) os preços se movem em tendências – o propósito de representar a ação dos preços em gráficos é identificar tendências, a partir dos estágios iniciais de desenvolvimento;
- c) a história se repete – os padrões comportamentais do ser humano tendem a se modificar de forma sutil ou permanecer invariável ao longo do tempo.

De acordo com a hipótese da eficiência perfeita dos mercados, seria impossível os investidores vencerem a economia, uma vez que os preços já incorporam todas as informações disponíveis para os resultados futuros. Isso significa que uma ação seria sempre transacionada no seu valor justo ou intrínseco, evitando lucros excessivos na compra de ações desvalorizadas ou venda de ações sobrevalorizadas (EATWELL; MILGATE; NEWMAN, 1994). O fato é que no mercado de ações ocorre enorme assimetria de informações; por exemplo: a diretoria de uma empresa sabe muito mais sobre ela do que os seus acionistas.

Por causa desta assimetria informacional, fatores de análise fundamentalista que poderiam contribuir para um movimento de alta, eventualmente, não farão o preço de uma ação subir, uma vez que os investidores com grande volume de negociação podem ter feito uma escolha no sentido contrário e se empenharão em proteger sua posição. Da mesma forma, o preço de uma ação já poderia ter subido bastante em virtude da antecipação dos *insiders* (investidores mais bem informados), devido a uma notícia favorável a ser divulgada pela empresa, mas agora que essa informação é de domínio público, o preço “inexplicavelmente” começa a cair.

A importância da análise técnica se justifica pois o mercado reflete a soma de todos os medos, desejos e expectativas dos seus participantes, e tais manifestações podem ser observadas em padrões gráficos. Por exemplo: os investidores recordam valores históricos onde realizaram grandes lucros ou prejuízos, formando zonas de preços difíceis de ultrapassar, as chamadas regiões de suporte e resistência. As tendências são formadas de modo semelhante, e as ferramentas da análise técnica permitem avaliar tanto sua força quanto sua provável extensão (MURPHY, 1999).

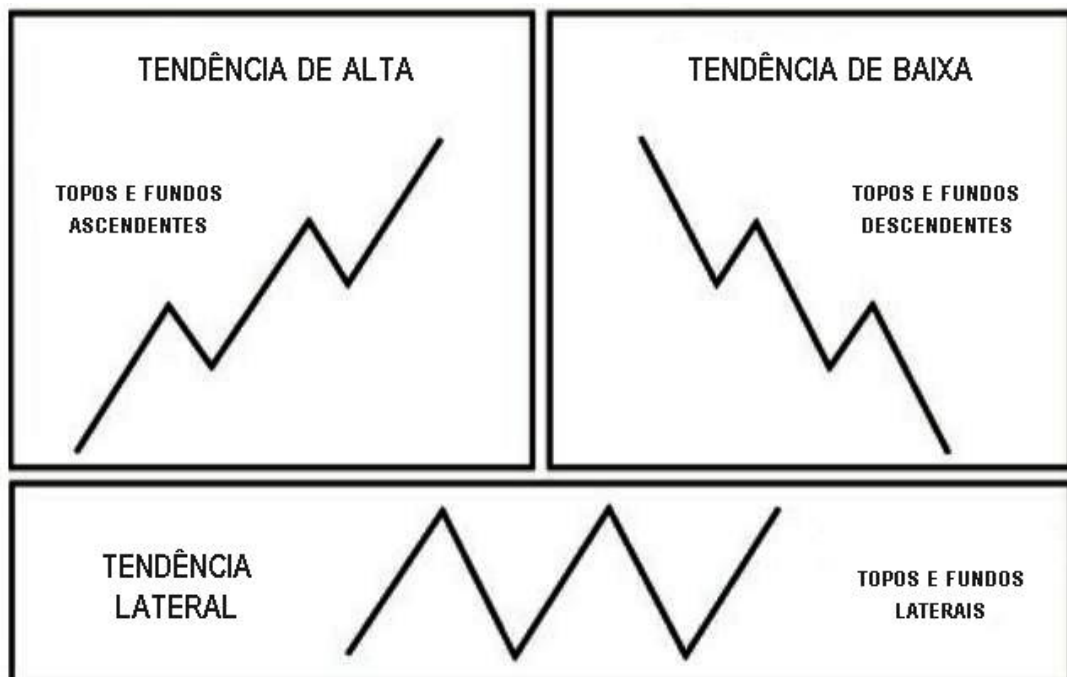
Uma das principais críticas à análise técnica seria a teoria *Random Walk*. Essa teoria preconiza a não utilização de tendências e padrões ocorridos no passado para

tentar prever o comportamento futuro do mercado, pois o movimento dos preços seria irracional e imprevisível, como se fosse um “passeio aleatório”. Entretanto, no final da década de 1980, o *Massachusetts Institute of Technology* (MIT) publicou um estudo do mercado acionário norte-americano, do período de 1962 a 1985, e provou que o movimento das ações não é randômico (LO; MACKINLAY, 1988).

2.2 Tendências de Mercado

A tendência dos preços é a direção para qual o mercado está se movendo. Um dos objetivos essenciais da análise técnica é identificar uma tendência em seu estágio inicial, de modo que as operações possam acompanhar a maior parte do movimento. As tendências podem apresentar três direções: para cima, para baixo e para o lado (esta última também chamada de congestão ou lateralidade), as quais podem ser observadas pelo sentido dos topos e fundos gerados graficamente (BROOKS, 2012). A Figura 4 apresenta estas características.

Figura 4 – Direções das tendências de mercado



Fonte: Elaborado pelo autor (2022)

Em uma tendência de alta, os preços alcançam topos e fundos ascendentes, enquanto numa tendência de baixa, os preços realizam topos e fundos descendentes. Já numa zona de congestão, os preços atingem topos e fundos em um mesmo nível, ou seja, ficam oscilando dentro de uma faixa relativamente limitada.

As tendências também são classificadas em primária, secundária e terciária. A tendência primária, também chamada de principal, representa movimentos de longo prazo e, geralmente, duram de três a 12 meses. A secundária ou intermediária, define movimentos de médio prazo, com períodos que variam entre um e três meses, e sinalizam correções da tendência principal. E, por fim, a terciária ou de curto prazo, com duração inferior a quatro semanas, que reforça ou contraria a tendência principal (BROOKS, 2012). Essa classificação das tendências pode ser observada através do uso de três médias móveis – de curto, médio e longo prazos, conforme a Figura 5.

Figura 5 – Identificação das tendências primária, secundária e terciária



Fonte: StockCharts.com (2021)

Outro aspecto importante: o volume é que move o preço, ou seja, a quantidade de ações negociadas deve acompanhar a variação de preços. Quando os preços se movimentam, o volume deve expandir na direção da tendência e contrair no

movimento de correção. Preços estáveis por algum tempo são um sinal que a relação entre a oferta e a demanda de ações está equilibrada. Vale ressaltar que a direção do mercado é sempre dada pelo movimento dos preços, o chamado *price action*, e a análise do volume funciona como uma confirmação do movimento (BROOKS, 2012).

2.3 Indicadores Financeiros

Os indicadores são importantes ferramentas da análise técnica para a previsão de mudanças de comportamento de um ativo, e utilizam informações sobre o preço e/ou volume. A formação de padrões gráficos nem sempre é tão clara; aliás, algumas vezes se torna ambígua ou de difícil interpretação. Diante disso, as principais funções dos indicadores são: reduzir a subjetividade da análise dos preços, antecipar a formação de tendências e detectar pontos de divergência (COLBY, 2010).

Um indicador técnico é calculado através de uma abordagem matemática tendo como entrada os preços do ativo – abertura, máxima, mínima ou fechamento. Alguns também consideram o volume negociado e o número de contratos em aberto. De fato, os indicadores proporcionam uma fonte extremamente útil de informações adicionais, em conjunto com os padrões gráficos e o próprio movimento dos preços. Eles ajudam a identificar a impulsão, a volatilidade e as divergências para, principalmente, auxiliar na análise técnica de tendências (MURPHY, 1999). Basicamente, os indicadores são classificados em duas categorias:

- a) Rastreadores – são aqueles que acompanham de perto os movimentos de mercado, mas nunca lideram o preço do ativo, por isso são chamados de indicadores atrasados. Funcionam melhor quando o mercado se encontra em tendência forte ou moderada, mas não são eficazes em mercados laterais.
- b) Osciladores – chamados de indicadores antecedentes, pois são projetados para antecipar os movimentos de preço e são mais adequados quando não houver uma tendência definida. Trabalham dentro de uma faixa de variação e sinalizam períodos de sobrecompra ou sobrevenda, além de auxiliarem na identificação de divergências.

Vale ressaltar também que alguns indicadores são construídos com base no volume, dada a premissa de que este precede o preço, já que os *insiders* procuram acumular ou distribuir antes do ativo entrar em tendência. Além disso, quanto maior o

volume negociado, maior a liquidez; e, em geral, quanto maior a liquidez, menor a volatilidade dos preços. Os indicadores de volume são utilizados para determinar a saúde de uma tendência e estimar sua intensidade. A falta de confirmação destes sinais pode indicar uma eventual reversão do movimento.

Além da classificação em rastreadores e osciladores, existem aqueles que se enquadram em classes menos triviais, como indicadores de sentimento, de fôlego de mercado, entre outros. De fato, existem centenas de indicadores e, todo ano, são criados novos deles. Mas somente aqueles que resistiram ao teste de longevidade são dignos de atenção e oferecem uma boa perspectiva. Tentativas de acompanhar muitos indicadores ao mesmo tempo se tornam inúteis, sendo melhor se concentrar em apenas dois ou três que se complementam (COLBY, 2010).

Como dito anteriormente, os indicadores são determinados a partir de cálculos matemáticos que são baseados em informações disponíveis de preço e volume. Desta forma, tais cálculos só poderão ser feitos sobre dados existentes, ou seja, que já fazem parte do passado – mesmo que muito recente. Por isso, é razoável considerar que qualquer indicador, seja rastreador ou oscilador, estará sempre “atrasado” em relação ao movimento dos preços.

A seguir serão apresentados, de forma sucinta, os indicadores utilizados neste trabalho, dando ênfase às características operacionais em detrimento às equações utilizadas para seus respectivos cálculos.

2.3.1 Rastreadores

- *Bollinger Bands (BB)*: esse indicador foi desenvolvido por John Bollinger, em 1980, e utiliza uma média móvel de 20 períodos, a partir da qual, são adicionadas duas bandas, uma superior e uma inferior, que se situam, respectivamente, a dois desvios-padrão desta média móvel. Quando o mercado entra em tendência, de alta ou de baixa, inclinando a média móvel para cima ou para baixo, respectivamente, as bandas se “abrem”, ou seja, se distanciam entre si. Durante esses movimentos, quando o preço do ativo cruza uma das bandas, tende a ocorrer um retorno à média móvel, indicando uma possível correção na tendência. Já em mercados laterais, as bandas superior e inferior se “fecham”, ou seja, vão se aproximando entre si. Este comportamento das bandas de Bollinger pode ser observado na Figura 6.

Figura 6 – Bandas de Bollinger BB



Fonte: StockCharts.com (2021)

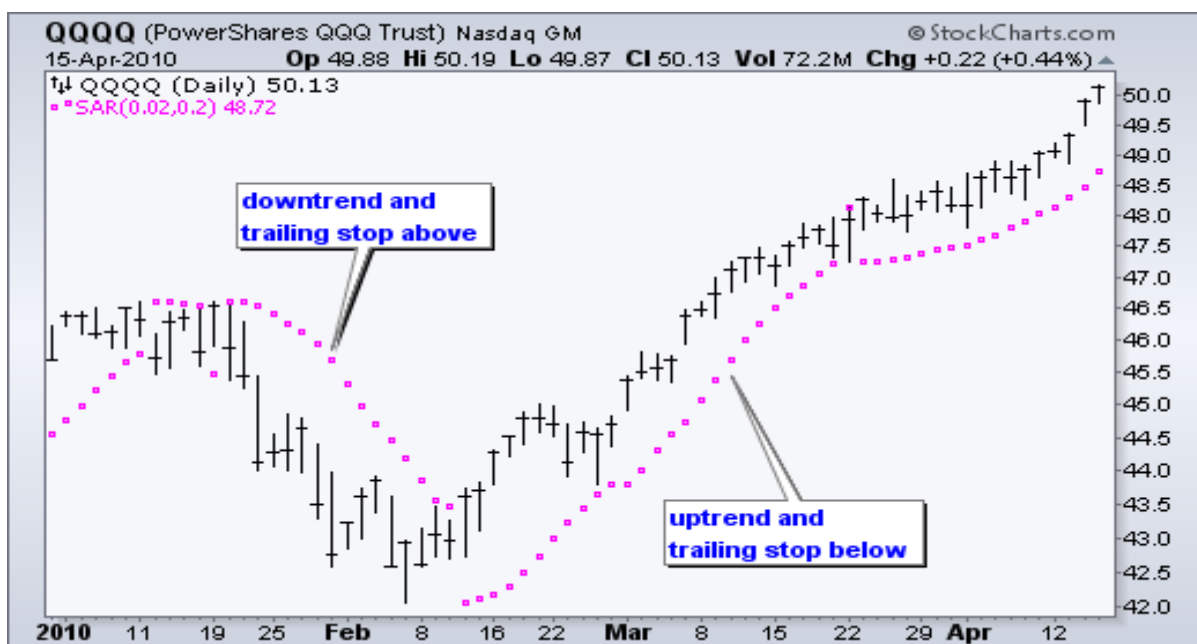
- *Average Directional Movement Index (ADX)*: criado por J. Welles Wilder, em 1978, o índice de movimento direcional é um dos indicadores mais usados para medir a força e a direção de uma tendência. Apresenta três componentes: o +DI (índice direcional positivo), o -DI (índice direcional negativo) e o ADX, que é resultado da combinação dos anteriores. O componente será positivo (+DI) quando a máxima atual do preço menos a máxima anterior for maior que a mínima anterior do preço menos a mínima atual, desde que seja positivo. Quando ocorrer um valor negativo é tratado como zero. O componente será negativo (-DI) quando a máxima atual do preço menos a máxima anterior for menor que a mínima anterior do preço menos a mínima atual, desde que seja positivo. Quando ocorrer um valor negativo é tratado como zero. Já o componente ADX, é igual ao valor absoluto da diferença entre +DI e -DI dividido pela soma de +DI e -DI. Valores do ADX acima de 20 confirmam entrada de tendência. Já o cruzamento entre as linhas +DI e -DI sinaliza a direção do movimento: se +DI estiver acima de -DI, temos mercado de alta; caso contrário, se -DI estiver acima de +DI, temos mercado de baixa. As características do indicador ADX, posicionado abaixo do gráfico de preços, estão na Figura 7.

Figura 7 – Índice de Movimento Direcional ADX



- *Parabolic SAR*: mais um rastreador de tendência desenvolvido por J. Welles Wilder, no final na década de 1970, onde SAR significa *stop and reverse*, que no jargão financeiro pode ser traduzido como “zere e vire a mão”. O indicador fica abaixo dos preços quando o mercado se encontra em tendência de alta; se o indicador ficar acima dos preços, temos um mercado em tendência de baixa. Se uma tendência permanecer forte, o parabólico SAR ganha incrementos de aceleração, ou seja, aumenta-se o espaçamento entre os pontos do indicador. No momento que a tendência perde força, o indicador vai se aproximando do preço, até chegar ao ponto em que ocorre a inversão. O SAR é também uma das ferramentas mais eficientes de *trailing stop* (proteção de lucros). A Figura 8 ilustra este indicador.

Figura 8 – Parabólico SAR



Fonte: StockCharts.com (2021)

2.3.2 Osciladores

- Moving Average Convergence/Divergence (MACD):** este oscilador desenvolvido no final dos anos de 1970, por Gerald Appel, é capaz de seguir tendências e avaliar impulsão. Consiste na diferença de duas médias, uma rápida de 12 períodos e outra lenta de 26 períodos, que convergem e divergem entre si. Desta forma, o indicador MACD flutua acima e abaixo de uma linha de zero, assumindo valores positivos quando a média rápida está acima da lenta, nos casos de tendência de alta, e valores negativos em tendência de baixa, quando a média rápida fica abaixo da lenta. Além dessa diferença entre médias, que constitui a linha MACD, também é calculada uma média de 9 períodos sobre o próprio MACD, chamada de linha de sinal, e quando ambas se cruzam sinalizam possíveis reversões na tendência. Outra importante característica deste oscilador é sua capacidade de capturar divergências em relação ao gráfico de preços. A Figura 9 ilustra este indicador.

Figura 9 – Média Móveis Convergentes / Divergentes MACD



Fonte: StockChars.com (2021)

- *Commodity Channel Index (CCI)*: é um oscilador que foi descrito em 1980, por Donald Lambert, para identificar os movimentos cíclicos nas *commodities*, mas que pode ser aplicado para qualquer mercado – ações, índices, etc. É utilizado tanto para detectar o início ou fim de uma tendência, quanto para identificar regiões de sobrecompra e sobrevenda dos ativos. Seu cálculo utiliza o preço típico (média da máxima, mínima e fechamento), uma média móvel de 20 períodos sobre o preço típico, além do desvio-padrão. Trata-se de um oscilador em bandas, sem limites definidos, que flutua a maior parte do tempo entre +100 e -100. Uma tendência de alta é sinalizada quando o oscilador cruza para cima a linha central, enquanto uma tendência de baixa é detectada quando cruza para baixo o nível de zero. Níveis extremos de +200 e -200 são muito mais difíceis de se alcançar, representando fortes regiões de sobrecompra e sobrevenda, respectivamente, e alertando para potenciais movimentos de reversão de tendência. A Figura 10 ilustra o CCI.

Figura 10 – Índice do Canal de Commodities CCI



Fonte: StockCharts.com (2021)

- *Relative Strength Index (RSI)*: outro indicador desenvolvido por J. Welles Wilder, em 1978, o Índice de Força Relativa compara a magnitude dos recentes ganhos do ativo com a magnitude de suas perdas, num período de 14 dias, e transforma essa informação em um número que vai de 0 a 100, ou seja, é um oscilador com bandas limitadas. Geralmente, tendências de alta se formam em movimentos ascendentes do indicador cruzando acima de 30, enquanto tendências de baixa se desenvolvem no cruzamento para baixo de 70. Quando o RSI se aproxima de 100, significa que a média das perdas é próxima de zero e que, nos últimos 14 dias, o ativo fechou seguidamente em alta, atingindo uma área sobrecomprada. De maneira análoga, se o RSI está próximo de zero, então a média das altas é quase nula e o fechamento do ativo está continuamente em baixa pelas últimas semanas, levando o ativo à região de sobrevenda. As características operacionais do Índice de Força Relativa (IFR) estão apresentadas na Figura 11.

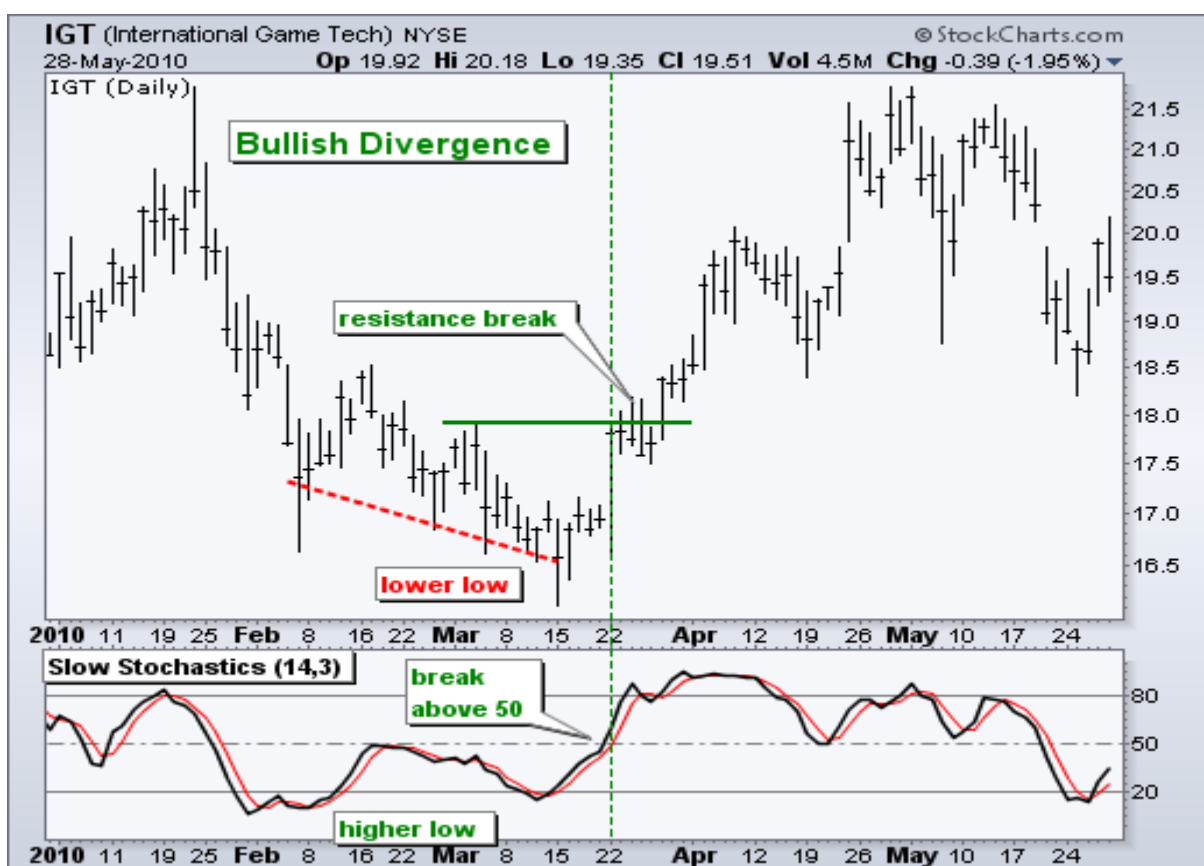
Figura 11 – Índice de Força Relativa RSI



Fonte: StockCharts.com (2021)

- *Stochastic Oscillator (STO)*: é um dos indicadores mais antigos, desenvolvido por Gerald Lane nos anos 1950. Também é um oscilador com bandas limitadas, que varia entre 0 e 100, e sua característica é calcular a posição em que o preço está com relação à amplitude das máximas e mínimas do ativo, num período de 14 dias. Este indicador é composto por duas linhas principais: a %K, que é a linha de sinal do oscilador e referência do preço em relação às máximas e mínimas; e a %D, que é uma média suavizada de 3 dias do %K. Geralmente, quando a %K cruza para cima a faixa central de 50 sinaliza tendência de alta e, quando cruza para baixo este nível, indica tendência de baixa. Na região de sobrecompra, acima de 80, se a %K cruza para baixo a %D é um forte sinal para reversão de baixa; enquanto na região de sobrevenda, abaixo de 20, se %K cruzar para cima a %D é uma possível reversão de alta. Como outros osciladores, o estocástico é eficiente para capturar divergências. Um exemplo operacional deste indicador está na Figura 12.

Figura 12 – Oscilador Estocástico STO



Fonte: StockCharts.com (2021)

2.3.3 Indicadores de Volume

- *On-Balance Volume (OBV)*: foi um dos primeiros indicadores de medida do fluxo de dinheiro, desenvolvido por Joe Granville em 1963. É um índice cumulativo que adiciona o volume dos dias de alta e subtrai o volume dos dias de baixa, ou seja, quando o preço de fechamento for acima do anterior o volume negociado é positivo; quando o fechamento for inferior ao anterior, o volume é negativo. Se o fechamento for igual ao dia anterior, então o indicador não sofre alteração. Dessa forma, um índice OBV crescente significa uma pressão de volume positiva, antecipando uma eventual tendência de alta. Por outro lado, um movimento decrescente no índice OBV pode sinalizar a formação de uma possível tendência de baixa. O indicador *On-Balance Volume* está ilustrado abaixo na Figura 13.

Figura 13 – On-Balance Volume OBV



- *Accumulation / Distribution Line (A/D)*: é um indicador cumulativo de volume criado por Marc Chaikin, em 1975, também com o objetivo de medir o fluxo de dinheiro. A principal diferença do OBV é que este, além do preço de fechamento, leva em consideração os preços de abertura, máxima e mínima do ativo, para criar um fator multiplicativo. Dessa forma, quanto mais próximo o preço de fechamento estiver da alta do dia, mais volume será adicionado ao total acumulado. Em contrapartida, quanto mais próximo o preço de fechamento estiver da mínima do dia, mais volume será subtraído do total acumulado. Se o preço de fechamento permanecer o mesmo, o total acumulado não se altera. O indicador A/D é usado para avaliar as tendências de preços e, potencialmente, identificar reversões futuras. Por exemplo, se o preço de um ativo se movimenta para baixo enquanto a linha A/D estiver subindo, pode indicar pressão compradora e o mercado se reverter para cima. A Figura 14 apresenta o comportamento deste indicador.

Figura 14 – Linha de Acumulação / Distribuição A/D



Fonte: StockCharts.com (2021)

- *Volume Weighted Average Price (VWAP)*: é uma média móvel ponderada pelo volume financeiro de um ativo. Diferente de outras médias móveis que dão peso ao tempo, a VWAP valoriza o volume negociado, indicando os movimentos de grandes *players* do mercado. Instituições financeiras e investidores estrangeiros utilizam a VWAP como *benchmark* para avaliação da eficiência na execução de suas ordens. Por tratar-se de um nível de preços com maior quantidade de negócios, devido ao rebalanceamento de posições compradas e vendidas, a VWAP funciona como uma região dinâmica de suporte e resistência. Preços se movimentando acima deste indicador demonstra tendência de alta; no caso contrário, preços cruzando abaixo da linha da VWAP pode indicar a formação de tendência de baixa. Este indicador está ilustrado na Figura 15.

Figura 15 – Preço médio ponderado pelo volume VWAP



Fonte: StockCharts.com (2021)

Capítulo 3

Árvores de Decisão, Random Forests e Desempenho

Neste capítulo serão abordados conceitos e características importantes tanto das árvores de decisão quanto dos algoritmos *Random Forests*. Além disso, também serão descritas a validação cruzada, a matriz de confusão e as métricas de desempenho mais utilizadas atualmente.

3.1 Árvores de Decisão

Uma árvore de decisão é um algoritmo de aprendizado supervisionado usado para classificação, bem como em problemas de regressão. Trata-se de um fluxograma com estrutura de uma árvore “invertida”, utilizado para prever resultados a partir de divisões baseadas em atributos. Esse processo de subdivisões, com base num teste de valor para cada atributo do conjunto de dados, é chamado particionamento recursivo (BREIMAN et al., 1998).

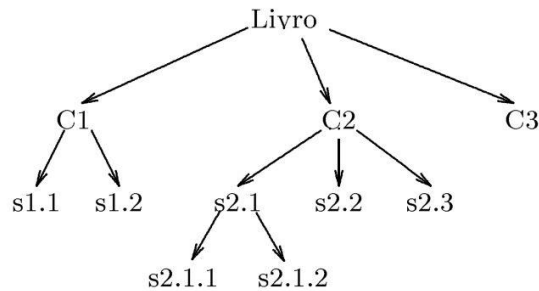
A árvore de decisão é uma coleção de elementos chamados nós, dentre os quais um é distinguido como o nó-raiz, que impõe uma estrutura hierárquica sobre todos os outros nós (LAURETTO, 1996). Formalmente, temos o seguinte:

- a) um único nó é uma árvore, sendo que este nó também é a raiz da árvore;
- b) suponha que t seja um nó e T_1, T_2, \dots, T_k sejam árvores com raízes t_1, t_2, \dots, t_k , respectivamente. Podemos construir uma nova árvore transformando t no pai dos nós t_1, t_2, \dots, t_k . Nessa árvore, t será a raiz e T_1, T_2, \dots, T_k serão as subárvores ou ramos da raiz. Os nós t_1, t_2, \dots, t_k são chamados filhos do nó t .

A Figura 16 ilustra o sumário de um livro representado através de uma árvore. A raiz, chamada Livro, possui três subárvores cujas raízes correspondem aos capítulos C1, C2 e C3, os quais são nós-filhos de Livro. Se t_1, t_2, \dots, t_k é uma sequência de nós em uma árvore tais que t_i é o pai de t_{i+1} para $1 \leq i \leq k$, então esta sequência é denominada um caminho do nó t_1 até o nó t_k . No exemplo da Figura 16, a sequência C2, s2.1, s2.1.2 é um caminho do capítulo C2 até a seção s2.1.2. Todos os nós que

não possuem filhos são chamados de nós-terminais ou folhas, tais como: C3 e s2.3. Já os nós que contém filhos são chamados de nós internos ou não terminais, tais como: C1 e s2.1 (LAURETTO, 1996).

Figura 16 – Representação de um sumário de livro numa árvore de decisão



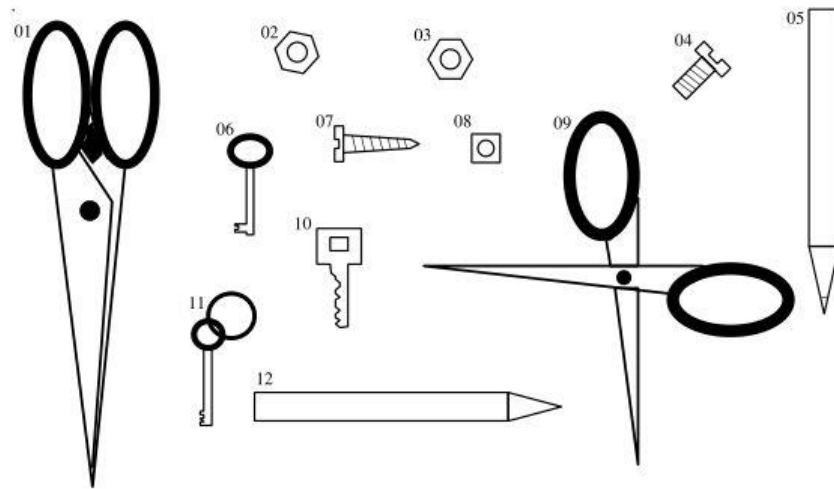
Fonte: Adaptado de LAURETTO (1996)

3.1.1 Construção de uma árvore de decisão

Para ilustrar a construção de uma árvore será utilizada a Figura 17, que mostra um conjunto de treinamento E, com objetos numerados de 01 a 12, e que pertencem a cinco classes: chave, porca, parafuso, caneta e tesoura. Suponha que as silhuetas sejam capturadas por uma câmera e processadas por um sistema de visão, capaz de extrair atributos de cada imagem. Os atributos são: o tamanho, o formato e o número de orifícios do objeto (LAURETTO, 1996). Os possíveis valores desses atributos são:

- Tamanho – pequeno, grande
- Formato – longo, compacto, outros
- Orifícios – 0, 1, 2, 3, muitos

Figura 17 – Conjunto E de imagens de objetos capturados por câmera



Fonte: Adaptado de LAURETTO (1996)

Assumindo que o sistema de visão tenha extraído os valores dos três atributos de cada objeto, a Tabela 1 mostra os respectivos vetores de atributos.

Tabela 1 – Vetores de atributos dos objetos do conjunto E

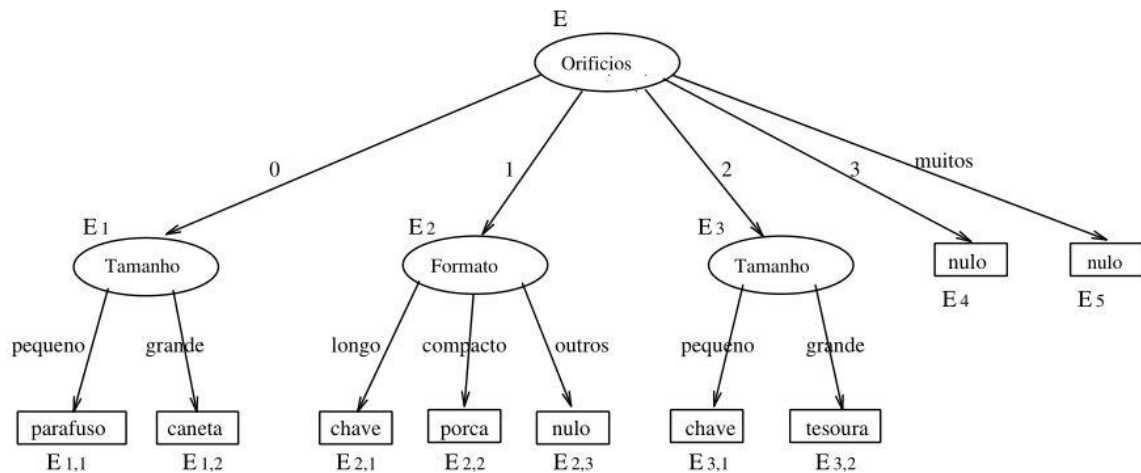
Objeto nº	TAMANHO	FORMATO	ORIFÍCIOS	CLASSE
1	grande	longo	2	tesoura
2	pequeno	compacto	1	porca
3	pequeno	compacto	1	porca
4	pequeno	longo	0	parafuso
5	grande	longo	0	caneta
6	pequeno	longo	1	chave
7	pequeno	longo	0	parafuso
8	pequeno	compacto	1	porca
9	grande	longo	2	tesoura
10	pequeno	longo	1	chave
11	pequeno	longo	2	chave
12	grande	longo	0	caneta

Fonte: Adaptado de LAURETTO (1996)

Para a construção da árvore, um atributo é escolhido para rotular o nó-raiz, por exemplo: Orifícios. Em seguida, o conjunto E é dividido em subconjuntos disjuntos, agrupando exemplos de mesmo valor de orifícios em cada subconjunto. Para cada

um dos subconjuntos, um nó-filho é criado a partir do nó Orifícios. A figura 18 mostra uma árvore de decisão induzida a partir do conjunto de treinamento E:

Figura 18 – Árvore de decisão induzida a partir do conjunto E



Fonte: Adaptado de Lauretto (1996)

Para a árvore de decisão induzida na Figura 18, os subconjuntos de E são:

$$E_1 = \{X \in E \text{ tal que } X_{\text{Orifícios}} = 0\}$$

$$E_2 = \{X \in E \text{ tal que } X_{\text{Orifícios}} = 1\}$$

$$E_3 = \{X \in E \text{ tal que } X_{\text{Orifícios}} = 2\}$$

$$E_4 = \{X \in E \text{ tal que } X_{\text{Orifícios}} = 3\}$$

$$E_5 = \{X \in E \text{ tal que } X_{\text{Orifícios}} = \text{muitos}\}$$

Uma vez que o subconjunto E_1 contém exemplos de classes distintas, o nó pode ser dividido, e um novo atributo é rotulado – Tamanho, e novos filhos são criados para este nó, conforme as características de tamanho (pequeno ou grande), obtendo os seguintes subconjuntos:

$$E_{1,1} = \{X \in E \text{ tal que } X_{\text{Tamanho}} = \text{pequeno}\}$$

$$E_{1,2} = \{X \in E \text{ tal que } X_{\text{Tamanho}} = \text{grande}\}$$

Como os exemplos em $E_{1,1}$ pertencem à mesma classe, rotulamos o nó com a respectiva classe à qual os exemplos pertencem (no caso, parafuso), e o mesmo é

declarado como nó-terminal ou folha. O mesmo ocorre com o nó referente ao subconjunto $E_{1,2}$, cujos exemplos são todos da classe caneta. O procedimento é análogo para os demais subconjuntos E_2 , E_3 , E_4 e E_5 ; quando o subconjunto é vazio, como ocorre em E_4 , E_5 e $E_{2,3}$, o nó correspondente é rotulado como nulo.

Suponha que seja apresentado à esta árvore de decisão um vetor de atributos de um objeto de classe desconhecida, a fim de que o modelo classifique este novo objeto. O procedimento é realizar o teste de atributo na raiz, ir para o ramo correspondente, realizar o teste de atributo no segundo nó, ir para o próximo ramo, sucessivamente, até atingir um nó-terminal ou folha e ser rotulado e classificado. Por exemplo: um novo objeto que apresente o vetor de atributos,

(Tamanho = grande, Formato = longo, Orifícios = zero)

se for submetido à árvore construída a partir do conjunto de treino E , atingirá o nó-terminal referente ao subconjunto $E_{1,2}$ e, portanto, será classificado como uma caneta.

Na construção da árvore de decisão, apresentada da figura 18, tanto o nó-raiz quanto os nós internos foram rotulados ao acaso. No entanto, a escolha do melhor atributo a ser utilizado em um determinado nó é feita por meio de critérios, que são baseados em medidas de informação (também chamadas de medidas de impureza).

3.1.2 Critérios para o particionamento dos nós

Os critérios de particionamento buscam dividir o conjunto de dados a partir de um nó-pai, escolhendo aquele atributo que minimize o grau de impureza dos nós-filhos. Em um determinado nó, a impureza será nula se todos os exemplos nele pertencerem à mesma classe; e a impureza será máxima se todas as classes estão igualmente presentes naquele nó (WEBB; COPSEY, 2011).

Entre as principais medidas de impureza estão a Entropia, o Índice de Gini e a Razão de Ganho de Informação. Antes de caracterizar cada uma dessas medidas, vamos adotar a seguinte notação:

- φ como sendo um conjunto de treinamento;
- K como sendo o número de classes;
- \mathcal{X} como sendo um espaço de atributos;
- γ como sendo um vetor de atributos;
- $n_{\bullet,t}$ como o número de elementos de φ incidentes sobre o nó t ;
- $n_{k,t}$ como o número de elementos de φ incidentes sobre o nó t , e pertencente à classe k ;
- $\pi_{k,t}$ como a probabilidade (desconhecida) de um elemento $\gamma \in \mathcal{X}$ incidente sobre o nó t , e pertencente à classe k ;
- $\hat{\pi}_{k,t}$ como um estimador pontual de $\pi_{k,t}$ representando a frequência relativa da classe k sobre o nó t , ou seja:

$$\hat{\pi}_{k,t} = \frac{n_{k,t}}{n_{\bullet,t}}$$

- a) Entropia: embora este conceito seja aplicado na física como uma grandeza termodinâmica, nos algoritmos de aprendizado supervisionado a entropia é uma medida de impureza ou incerteza de informação (FRIZZARINI, 2013). O cálculo da entropia em um nó de uma árvore de classificação é dado por:

$$\mathcal{E}(t) = -\sum_{k=1}^K \hat{\pi}_{k,t} \cdot \log_2(\hat{\pi}_{k,t})$$

O atributo que será selecionado é o que apresentar a maior diferença entre a entropia do nó-pai e a média ponderada das entropias dos nós-filhos. Essa diferença é conhecida por Ganho de Informação e sua expressão é dada por:

$$\Delta\mathcal{E}(t) = \mathcal{E}(t) - \sum_{z=1}^Z \frac{N_{t_z}}{N_t} \cdot \mathcal{E}(t_z)$$

onde Z é o número de nós-filhos, N_t é o número total de elementos do nó-pai e N_{t_z} é o número de elementos do nó-filho t_z .

- b) Índice de Gini: esta medida de impureza foi desenvolvida e proposta por Conrado Gini, em 1912, e é muito empregada em análises econômicas e sociais. Em algoritmos de árvores de classificação esse índice é utilizado como critério de particionamento dos nós (FRIZZARINI, 2013). Sua expressão é:

$$G(t) = 1 - \sum_{k=1}^K \hat{\pi}_{k,t}^2$$

Sob essa medida, o atributo selecionado é o que apresenta o maior decréscimo no índice Gini na divisão do nó, obtido por:

$$\Delta G(t) = G(t) - \sum_{z=1}^Z \frac{N_{t_z}}{N_t} \cdot G(t_z)$$

onde Z é o número de nós-filhos, N_t é o número total de elementos do nó-pai e N_{t_z} é o número de elementos do nó-filho t_z .

- c) Razão de Ganho de Informação: o critério de partição baseado apenas no Ganho de Informação tende a dar maior preferência para atributos com muitas divisões possíveis. A medida de Razão de Ganho de Informação foi proposta para contornar este problema e é, conceitualmente, o Ganho de Informação ponderado (FRIZZARINI, 2013), sendo representada por $GR(t)$, cuja expressão é dada por:

$$SP(t) = - \sum_{z=1}^Z \frac{N_{t_z}}{N_t} \cdot \log_2 \left(\frac{N_{t_z}}{N_t} \right)$$

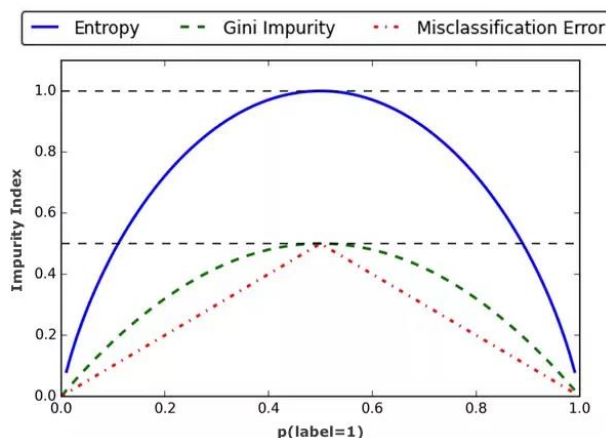
$$GR(t) = \frac{\Delta \mathcal{E}}{SP(t)}$$

onde Z é o número de nós-filhos, N_t é o número total de elementos do nó-pai e N_{t_z} é o número de elementos do nó-filho t_z .

A Figura 19 ilustra a diferença entre o índice Gini, a Entropia e uma taxa de erro esperado, em um problema de classificação binária. O eixo horizontal representa a proporção de exemplos de cada uma das classes (denotada por p). As três medidas, independentemente da quantidade de exemplos resultantes em cada nó, priorizam nós-filhos cada vez mais puros, como pode ser observado nas extremidades de p (p próximo de 0 ou 1), onde as três medidas apresentam o melhor resultado. Também é

possível observar que, pela concavidade das curvas Gini e Entropia, essas duas medidas são mais sensíveis às variações de p nas extremidades do que no centro (p próximo de 0,5), enquanto na medida de erro essa taxa de variação é constante.

Figura 19 – Comparação entre medidas de impureza de um nó



Fonte: Medium.com (2021)

Normalmente, os conjuntos de dados no mundo real têm um grande número de atributos, ou variáveis, o que resulta num grande número de divisões e gera árvores mais complexas. A construção de árvores muito grandes, além de consumir mais tempo computacional, pode levar ao *overfitting* (sobreajuste dos dados). Diante disso, a árvore fornecerá uma precisão muito boa no conjunto de treinamento, mas baixa precisão no conjunto de teste. No entanto, é possível interromper o crescimento da árvore através do ajuste de alguns parâmetros, como o número máximo de atributos a serem considerados para divisão e a profundidade máxima da árvore (ROKACH; MAIMON, 2005).

Existe um outro método que contribui para evitar o *overfitting* que é a poda. Este processo ajuda a melhorar o desempenho da árvore de decisão cortando os nós que são pouco significativos e removendo os ramos de baixa importância (WITTEN et al., 2017). Existem, basicamente, duas formas de se realizar a poda:

- Pré-poda: paramos de crescer a árvore mais cedo, ou seja, podemos remover um nó de baixa importância durante o crescimento da árvore;
- Pós-poda: com a árvore já construída em sua profundidade, é possível começar a remover os nós com base na sua importância.

3.2 Random Forests

As *Random Forests* são algoritmos de aprendizado supervisionado, que fazem parte dos chamados métodos *ensemble*, ou seja, são construídos pela combinação de algoritmos mais básicos, mas que os tornam mais robustos. Tal característica está na capacidade de combinar um grande número de árvores de decisão, com diferentes respostas, em um único resultado. Em problemas de classificação, o resultado mais votado ou que mais se repete, será o escolhido; no caso dos problemas de regressão, é calculada a média dos valores fornecidos pelos modelos para obtenção do resultado final (BREIMAN, 2001).

O objetivo principal das *Random Forests* é agrupar conjuntos de preditores não necessariamente ótimos, ao invés de buscar otimizar cada preditor de uma única vez, como acontece nos modelos CART (*Classification and Regression Trees*). Como as árvores individuais são perturbadas aleatoriamente, a floresta se beneficia de uma exploração mais extensa do espaço total de atributos, permitindo obter um melhor desempenho preditivo (GENUER; POGGI, 2020).

Cada árvore dentro da *Random Forest* é construída, basicamente, da seguinte maneira (GENUER; POGGI, 2020):

- A partir de um conjunto de treino com N observações, é feita uma amostragem *bootstrap*, com reposição, também de N observações. Essa amostra será o conjunto de treino para o crescimento de uma árvore de decisão;
- Supondo M atributos de entrada, um número $m \ll M$ é especificado, de forma que em cada nó, m variáveis sejam selecionadas de M , de forma aleatória, e a melhor divisão entre m será utilizada em cada nó;
- Cada árvore individual é construída na maior extensão possível, ou seja, não há poda.

Os algoritmos *Random Forests* funcionam de forma bastante eficiente quando aplicados a grandes volumes de dados, podendo lidar com centenas de atributos de entrada. Além disso, apresentam alta precisão quando comparados a outros modelos. Também fornecem estimativas de quais atributos são mais importantes na regressão ou classificação e geram estimativas internas de erro, à medida em que se avança na sua construção (BREIMAN, 2001).

Quando em um determinado conjunto de treinamento são definidas as reamostragens do tipo *bootstrap*, para construção de cada árvore de decisão, cerca de um terço destas observações são deixadas de fora de cada amostra. Esses registros são chamados *Out-of-Bag* e são usados internamente pela *Random Forest*, para obter uma estimativa não viesada do erro de classificação, à medida em que outras árvores são adicionadas ao modelo (GENUER; POGGI, 2020).

Para estimar o grau de importância de cada atributo do conjunto de dados, o algoritmo de *Random Forest* realiza o seguinte processo (GENUER; POGGI, 2020):

- Após cada árvore ser construída na floresta é realizada uma validação interna com as observações *out-of-bag*, calculando-se a proporção de instâncias *out-of-bag* que foram classificadas corretamente;
- Em seguida, permuta-se os valores de uma das variáveis m , de forma aleatória, dentro das instâncias *out-of-bag*, e outra validação interna é realizada pelo modelo;
- Por fim, é calculada a diferença entre a proporção de instâncias *out-of-bag* classificadas incorretamente, após permutação da variável m , e a proporção de instâncias classificadas incorretamente dos dados *out-of-bag* intocados. A média dessa diferença nos fornece a pontuação de importância da variável m .

A taxa de erro das *Random Forests* depende, principalmente, de dois fatores: da correlação entre duas árvores quaisquer do modelo, onde o aumento da correlação aumenta o erro; e da força individual de cada árvore, ou seja, uma árvore com baixa taxa de erro é um classificador forte e, assim, aumentar a força de árvores individuais diminui a taxa de erro como um todo (GENUER; POGGI, 2020).

3.3 Desempenho dos Classificadores

3.3.1 Validação Cruzada

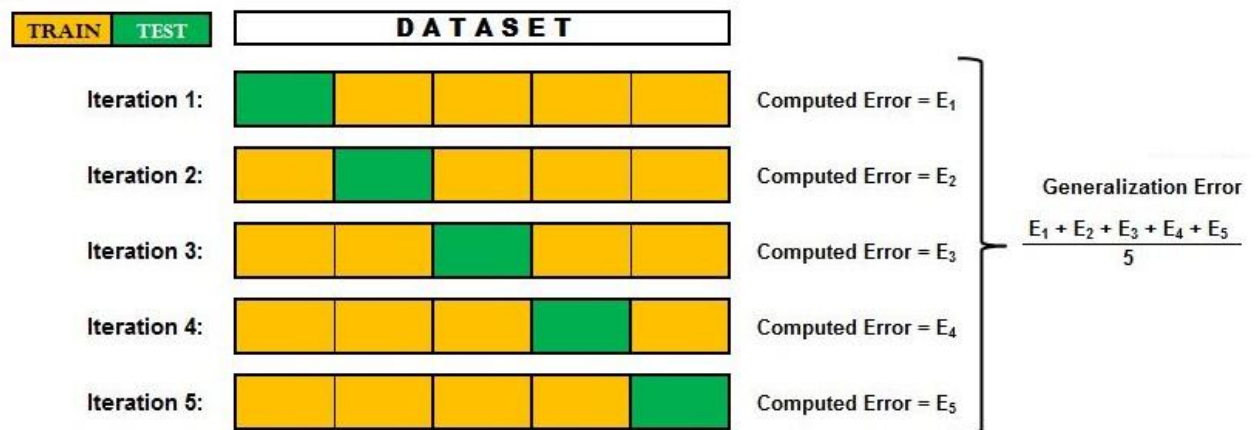
A validação cruzada é um método utilizado para estimar a precisão de um classificador de aprendizagem supervisionada. As várias técnicas de validação cruzada têm em comum o conceito de particionar o conjunto de dados em subconjuntos mutuamente exclusivos e, posteriormente, utilizar uma parte dos

subconjuntos para indução do algoritmo (conjunto de treinamento), e o restante para validação do modelo (conjunto de teste). Além de fornecer uma estimativa mais precisa do desempenho preditivo do algoritmo, a validação cruzada também reduz a ocorrência de *overfitting* nos dados (BERRAR, 2018).

Entre as principais técnicas de validação cruzada, podemos citar: o *k-Fold*, o *Leave-One-Out* e o *Holdout*.

- a) *k-Fold* – nesta técnica o conjunto de dados é dividido, aleatoriamente, em k subconjuntos de tamanhos iguais, sendo que em cada iteração um destes subconjuntos é deixado de fora, e o modelo é treinado com os outros $k-1$ subconjuntos. Em seguida, o algoritmo é validado no subconjunto isolado, e o processo é repetido k vezes, até que todos os subconjuntos sejam usados para validação do modelo. Ao final das iterações um erro médio é calculado, sendo que a variância da estimativa de erro diminui à medida que o número k de subconjuntos aumenta. A Figura 20 ilustra o *k-fold*, com k igual a 5, ou seja, o treinamento é realizado com 80% dos dados e 20% é para teste.

Figura 20 – Técnica *k-Fold* de validação cruzada



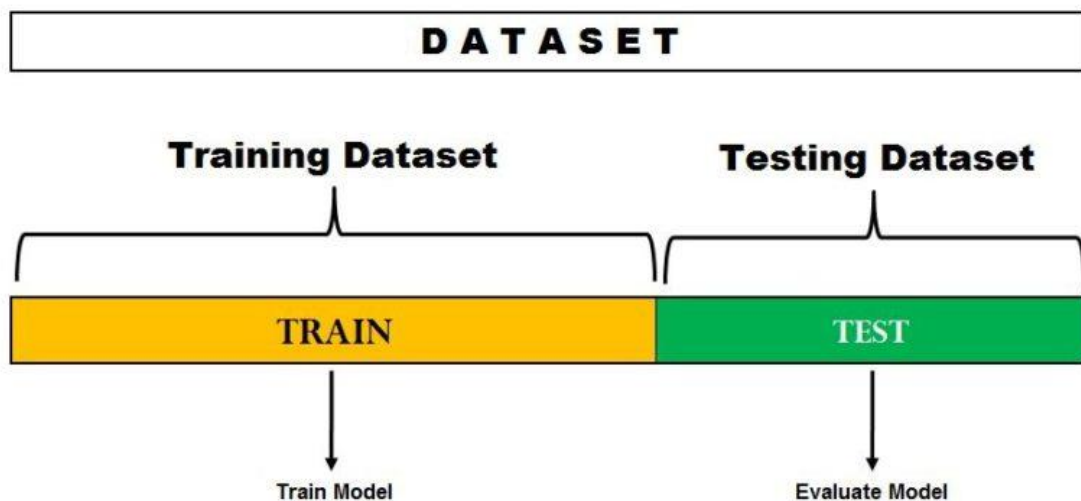
Fonte: DataVedas.com (2021)

- b) *Leave-One-Out* – esta técnica é um caso particular da validação *k-Fold*, onde K assume um valor igual ao número N de observações do conjunto de dados, ou seja, $K = N$. Dessa forma, cada instância será utilizada uma vez para teste e $N-1$ vezes para a indução do classificador. Como o processamento será realizado durante N iterações, o custo computacional é muito maior, por isso a

técnica *Leave-One-Out* é mais indicada quando o número de registros do conjunto de dados é pequeno;

- c) *Holdout* – é uma técnica bastante utilizada de validação cruzada, onde o conjunto de dados é dividido em dois subconjuntos, mutuamente exclusivos, sendo o primeiro utilizado para indução do classificador e o segundo para a validação. Geralmente, a divisão é feita na seguinte proporção: 2/3 dos dados para o subconjunto de treinamento (indução) e 1/3 para o subconjunto de teste (validação). A técnica *Holdout* com amostragem estratificada busca manter nos dois subconjuntos a mesma proporção entre as classes do conjunto original. O *Holdout* é recomendado para conjunto de dados com grande quantidade de exemplos, e foi a técnica utilizada neste trabalho. A Figura 21 ilustra a técnica.

Figura 21 – Técnica *Holdout* de validação cruzada



Fonte: DataVedas.com (2021)

3.3.2 Matriz de Confusão

Uma matriz de confusão é um método usado para medir o desempenho de um algoritmo de classificação. Ela pode ser aplicada a problemas de classificação binária, bem como para problemas de classificação multi-classe. O classificador é executado sobre o conjunto de testes que resulta em uma tabela, construída a partir da contagem (frequência absoluta) dos exemplos de validação. Essa matriz apresenta quatro características básicas, expressas em números, que depois serão utilizadas para

calcular as métricas de desempenho do classificador (GRANDINI; BAGLI; VISANI, 2020). Estas características básicas são apresentadas na Figura 22.

Figura 22 – Matriz de Confusão

		CLASSE PREDITA		
		P	N	
CLASSE REAL	P	VP Verdadeiro Positivo	FN Falso Negativo	POS Positivo total (real)
	N	FP Falso Positivo	VN Verdadeiro Negativo	NEG Negativo total (real)
		PP Positivo total (predito)	PN Negativo total (predito)	

Fonte: Elaborado pelo autor (2022)

sendo que,

- VP (Verdadeiros Positivos): quantidade de exemplos positivos que foram classificados corretamente;
- FN (Falsos Negativos): quantidade de exemplos positivos que foram classificados, erroneamente, como negativos;
- FP (Falsos Positivos): quantidade de exemplos negativos que foram classificados, erroneamente, como positivos;
- VN (Verdadeiros Negativos): quantidade de exemplos negativos que foram classificados corretamente.

Após as contagens, a matriz de confusão ainda fornece outras informações que consolidam as características obtidas:

- POS – é a quantidade real de exemplos positivos, ou seja, verdadeiros positivos somados aos falsos negativos: $POS = VP + FN$;
- NEG – é a quantidade real de exemplos negativos, ou seja, falsos positivos somados aos verdadeiros negativos: $NEG = FP + VN$;
- PP – é a quantidade predita de exemplos positivos, ou seja, verdadeiros positivos somados aos falsos positivos: $PP = VP + FP$;
- PN – é a quantidade predita de exemplos negativos, ou seja, verdadeiros negativos somados aos falsos negativos: $PN = VN + FN$.

3.3.3 Métricas de desempenho

A partir da matriz de confusão é possível calcular as taxas de erros e acertos do classificador, com o objetivo de encontrar as métricas de desempenho do modelo (GRANDINI; BAGLI; VISANI, 2020). As métricas utilizadas nesse estudo foram:

- a) Taxa de Precisão (VP / PP): é a proporção de exemplos positivos classificados corretamente em relação ao total de classificações positivas, ou seja, essa taxa corresponde à probabilidade estimada de um exemplo ser da classe positiva, dado que foi classificado como positivo;
- b) Taxa de Recall (VP / POS): também conhecida como sensibilidade, ela mede a capacidade do classificador atribuir corretamente a classe positiva. Trata-se da proporção de exemplos classificados corretamente como positivos em relação ao total real de casos positivos;
- c) Taxa de Acurácia (VP+VN / POS+NEG): é a proporção de classificações corretas em relação ao total de exemplos;
- d) F1-score: esse coeficiente relaciona as taxas de precisão e recall em um único número, através do cálculo de uma média harmônica. A métrica pode ser expressa como: $F1\text{-score} = 2 \times [(precisão \times recall) / (precisão + recall)]$. A partir do cálculo do F1-score, são calculadas: a métrica Macro-F1, que é a média aritmética entre os respectivos F1-score de cada uma das classes; e a métrica Ponderada-F1, que é a média ponderada entre os respectivos F1-score de cada classe, proporcional ao número de exemplos em cada uma delas;
- e) Correlação de Matthews: o coeficiente de Matthews é uma métrica que produz uma pontuação alta somente se a previsão do modelo obtiver bons resultados nas quatro categorias da matriz de confusão – VP, VN, FP e FN, proporcionalmente à quantidade de elementos positivos e negativos do conjunto de dados. Essa métrica varia entre -1 e 1, para erro total de classificação e classificação perfeita, respectivamente, enquanto 0 é o valor esperado para um classificador aleatório;
- f) Índice Kappa: esse coeficiente mede a dependência do acaso entre verdadeiros e preditos, ou seja, mede o grau de concordância entre os valores reais e os valores previstos. Essa métrica varia entre 1 e -1, considerando 1 como concordância perfeita e valores abaixo de 0 como concordância aleatória. Geralmente, se obtém um valor entre 0 e 1.

Capítulo 4

Métodos de Seleção de Atributos

Neste capítulo será apresentada a importância da seleção de atributos de um conjunto de dados, bem como uma descrição dos métodos de seleção que foram comparados – VSURF (*Variable Selection Using Random Forests*) e RFE (*Recursive Feature Elimination*). O método de seleção VSURF foi inteiramente descrito com base em (GENUER; POGGI, 2020), motivo pelo qual as citações foram suprimidas.

4.1 Importância da Seleção de Atributos

A engenharia de atributos é o processo de selecionar, manipular e transformar dados brutos em novas variáveis de entrada que, eventualmente, possam melhorar a capacidade do modelo de capturar as relações preditivas com a variável resposta. Entre as técnicas de engenharia de atributos está a expansão de base, que gera mais variáveis de entrada que o conjunto de dados original. O objetivo desta transformação é que os preditores criados no processo sejam ainda mais relevantes para o aprendizado de máquina (KUHN; JOHNSON, 2020).

Por outro lado, é possível que o desempenho preditivo seja degradado com o aumento de atributos pouco informativos, tornando necessário realizar uma seleção criteriosa das variáveis de entrada. Teoricamente, as principais motivações para se remover atributos de um conjunto de dados estão relacionadas com o aumento na precisão e menor tempo de processamento do algoritmo, bem como na redução da complexidade geral do modelo (KUHN; JOHNSON, 2020). Vale observar que:

- a regressão linear e a regressão logística são muito vulneráveis a preditores correlacionados, ou seja, a remoção de variáveis redundantes pode reduzir a multicolinearidade entre atributos, permitindo que tais modelos sejam melhor ajustados;
- as redes neurais artificiais e as máquinas vetoriais de suporte são modelos mais sensíveis a preditores irrelevantes, ou seja, a presença de

variáveis supérfluas causa uma redução significativa nos respectivos desempenhos preditivos;

- mesmo quando um modelo é insensível a preditores pouco informativos, é importante considerar o menor conjunto de variáveis possível, desde que forneça um resultado aceitável, uma vez que a redução de atributos influencia diretamente o custo computacional da modelagem.

De maneira geral, o desempenho de algoritmos preditivos, seja em aprendizado supervisionado ou não supervisionado, é diretamente influenciado pela capacidade de se criar atributos que capturem novos *insights* sobre o problema. Após esta engenharia, é feita uma seleção de quais preditores devem ser mantidos ou excluídos da base de dados (KUHN; JOHNSON, 2020). A Figura 23 apresenta uma classificação dos principais métodos de seleção de atributos.

Figura 23 – Métodos de seleção de atributos

Filter Method	Correlation, Chi-square test, Information gain
Wrapper Method	Forward Selection, Backward elimination, Stepwise selection
Embedded Method	Regularization methods such as LASSO, Ridge regression, Elastic net

Fonte: iZen Artificial Intelligence (2021)

Segue abaixo uma breve descrição de cada método:

- métodos tipo filtro: essa abordagem deriva da ideia que atributos irrelevantes sejam “filtrados” do conjunto de dados antes da aplicação do algoritmo. Podem avaliar cada atributo de forma isolada, determinando o grau de correlação entre cada preditor e a variável resposta, ou buscar subconjuntos de preditores que sejam mais eficientes na identificação das classes (BLUM; LANGLEY, 1997). Entre os métodos do tipo filtro, podemos citar: o *Information Gain Attribute Ranking* e a *Correlation-based Feature Selection*;

- b) métodos tipo *wrapper*: também chamados de métodos de encapsulamento, são técnicas de seleção de atributos que empregam algoritmos preditivos para avaliar os subconjuntos de variáveis de entrada. Esses modelos possuem critérios de parada e estratégias de busca configuráveis, sendo que a avaliação dos subconjuntos é feita com base na acurácia preditiva do algoritmo (LIU; MOTODA, 2007). Como exemplos de métodos *wrapper*, podemos citar as duas técnicas comparadas neste trabalho: a técnica VSURF – *Variable Selection Using Random Forests* e a técnica RFE – *Recursive Feature Selection*;
- c) métodos tipo *embedded*: também chamados de métodos incorporados, são técnicas que selecionam um subconjunto de atributos no próprio processo de construção do modelo, durante a fase de treinamento, e que possuem regras específicas para cada tipo de algoritmo utilizado (FONTI; BELITSER, 2017). Um exemplo típico dos métodos incorporados é a regressão do tipo LASSO, que possui funções de penalização embutidas para reduzir o *overfitting*.

A Tabela 2 apresenta um comparativo entre características importantes dos três tipos de métodos de seleção:

Tabela 2 – Comparação de características entre métodos de seleção

	Método FILTER	Método WRAPPER	Método EMBEDDED
O que é	usa medidas representativas	usa modelos preditivos	seleção de atributos está incorporada na construção do modelo
Velocidade	computacionalmente mais rápido	computacionalmente mais lento	intermediário
Overfitting	evita overfitting	propensão ao overfitting	menor propensão ao overfitting
Desempenho	algumas vezes pode falhar na seleção dos melhores atributos	excelente performance	boa performance

Fonte: Adaptado de iZen Artificial Intelligence (2021)

4.2 VSURF – Variable Selection Using Random Forest

O método VSURF é um processo automatizado, no sentido de que não é necessário especificar o número desejado de atributos, e que se adapta aos dados para fornecer um subconjunto de variáveis. O método consiste de duas etapas: a primeira, mais “grosseira” e decrescente, busca estabelecer o limite mínimo aceitável para a importância das variáveis e, assim, eliminar o maior número de atributos pouco informativos; a segunda etapa, mais refinada e ascendente, realiza uma introdução iterativa de variáveis que vai adicionando os atributos um de cada vez, num processo conhecido como *forward*.

Nesta segunda etapa do processo ocorrem duas fases distintas: interpretação e previsão. Na fase de interpretação são selecionados os atributos que estão fortemente relacionados com a variável resposta, mesmo que apresentem uma alta correlação entre si. A partir deste subconjunto resultante, a fase de previsão irá selecionar uma quantidade de variáveis suficiente para prever, de forma adequada, a variável de saída.

A estrutura geral do método VSURF pode ser descrita pelo seguinte processo:

a) Etapa 1 – Limiarização

- classificação dos atributos em ordem decrescente de importância através de um índice chamado *Variable Importance (VI)*, comentado na seção 3.3, que é calculado através da seguinte equação:

$$VI(x^j) = \frac{1}{q} \sum_{t=1}^q \left(\text{err}\widehat{O}\widehat{O}\widehat{B}_t^j - \text{err}OOB_t \right)$$

onde o índice de importância da variável $VI(x^j)$ é igual a diferença entre a proporção de instâncias *out-of-bag* classificadas incorretamente, após a permutação da variável x^j , e a proporção de instâncias classificadas incorretamente dos dados *out-of-bag* intocados. A média dessa diferença nos fornece a importância da variável x^j ;

- eliminação dos atributos de baixa importância utilizando um limite mínimo definido, chamado *threshold*.

Nesta primeira etapa, é considerada a sequência correspondente dos desvios-padrão dos índices de importância das variáveis, uma vez que a variabilidade dos VI 's é maior para as variáveis relevantes do que para variáveis pouco informativas. O *threshold* é definido por um valor mínimo de importância previsto por um modelo CART (*Classification And Regression Tree*). Apenas as variáveis cuja importância média VI for maior do que o *threshold* estabelecido é que serão selecionadas nesta fase de limiarização.

b) Etapa 2 – Interpretação e Predição

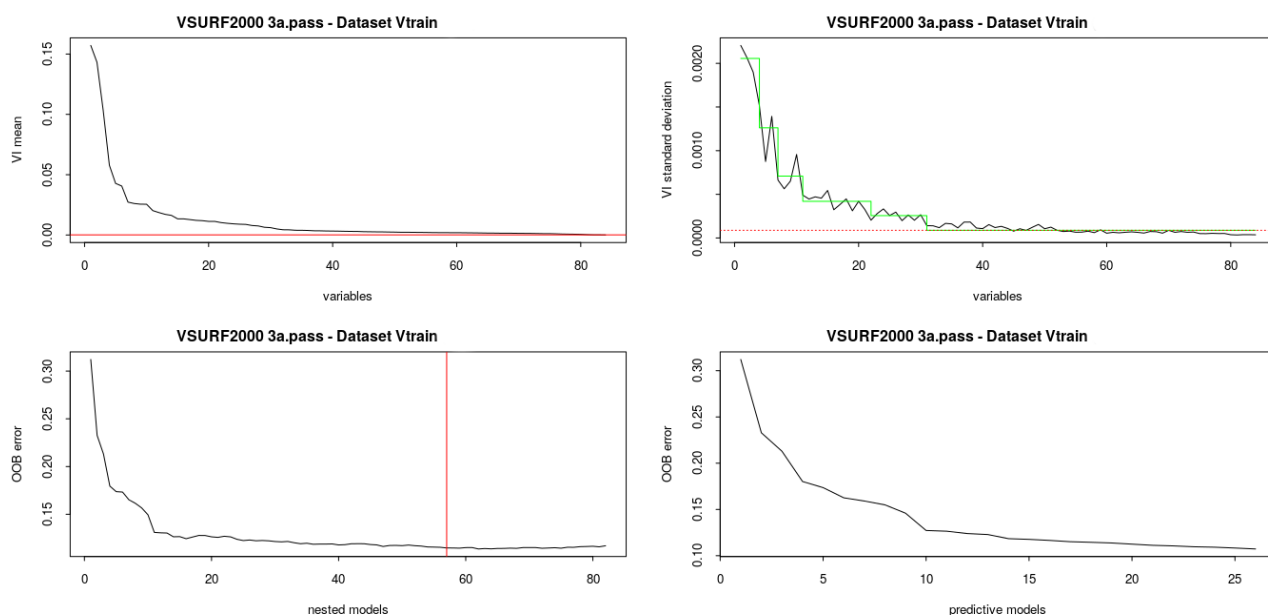
- *Fase Interpretação*: é construída uma coleção de modelos aninhados usando as primeiras k variáveis (onde k são as mais importantes), com $k = 1$ até m , sendo m o número de variáveis que passaram pela limiarização. Então, são selecionadas as variáveis que conduzem ao menor erro OOB, calculando as médias dos erros OOB dos modelos aninhados, começando com aquele que tem apenas a variável mais importante e terminando com aquele que envolve todas as m variáveis, previamente selecionadas na limiarização. Normalmente, as m' variáveis do modelo que levam ao menor erro OOB são selecionadas nesta fase;
- *Fase Predição*: a partir das variáveis selecionadas pela interpretação, uma sequência de subconjuntos é construída, introduzindo, sequencialmente, as variáveis em ordem crescente de importância e testando-as iterativamente. A introdução sequencial de variáveis é baseada na seguinte premissa: uma variável é adicionada apenas se o erro OOB diminuir mais do que um limite. A ideia é que o erro OOB deve diminuir mais do que a variação média gerada pela inclusão de variáveis não informativas. O limite é definido como a média dos valores absolutos das diferenças dos erros OOB entre os subconjuntos, incluindo o de m' variáveis e aquele com m variáveis, conforme a equação abaixo:

$$\frac{1}{m - m'} \sum_{j=m'}^{m-1} |errOOB(j+1) - errOOB(j)|$$

onde $errOOB(j)$ é o erro *out-of-bag* da floresta construída com as j variáveis mais importantes.

Os gráficos apresentados na Figura 24 ilustram o processo VSURF de seleção de variáveis em suas respectivas etapas. As duas janelas gráficas na parte superior representam a primeira etapa descrita do processo, ou seja, a limiarização. Já os gráficos na parte inferior ilustram a segunda etapa: fase de interpretação na esquerda e a fase de predição na direita, respectivamente.

Figura 24 – Etapas do método de seleção de variáveis VSURF



Fonte: Elaborado pelo autor (2022)

4.3 RFE – Recursive Feature Elimination

O método RFE, ou Eliminação de Atributos Recursiva, é uma seleção do tipo *backward*, ou seja, começa a construir o modelo com o conjunto total de preditores e pontua a importância de cada atributo. As variáveis de entrada menos importantes são então removidas, o modelo é reconstruído e as pontuações de importância são computadas novamente, até que um subconjunto ótimo seja selecionado. (KUHN; JOHNSON, 2016).

Nem todos os modelos são adequados para trabalhar em conjunto com o RFE, e alguns algoritmos se beneficiam mais deste método de seleção de variáveis do que outros. A principal desvantagem do RFE é que ele requer que o modelo inicial seja capaz de ajustar todo o conjunto de preditores, sendo que alguns algoritmos não podem ser utilizados quando o número de preditores for maior que o número de

observações, tais como a regressão linear múltipla, a regressão logística e a análise discriminante linear (KUHN; JOHNSON, 2016).

Por outro lado, as *Random Forests* são algoritmos que trabalham de forma eficiente com o método RFE, pelos seguintes motivos (SVETNIK et al., 2003):

- a) as *Random Forests* tendem a não excluir variáveis da função de predição. O motivo está relacionado à natureza dos subconjuntos do modelo, ou seja, o aumento do desempenho em métodos *ensemble* está relacionado à diversidade nos modelos constituintes, e a média dos subconjuntos que são efetivamente os mesmos não reduz a variação nas previsões do algoritmo. Por esta razão, as *Random Forest* forçam as árvores a conter divisões sub-ótimas dos preditores usando uma amostra aleatória de todas as variáveis de entrada;
- b) as *Random Forests* são algoritmos que possuem um método interno para medir a importância de atributos, e utilizam todos os preditores do conjunto de dados para calcular as classificações iniciais de importância, conforme comentado na seção 3.3. Mesmo as variáveis de entrada altamente correlacionadas, mas que sejam úteis na previsão do modelo, são utilizadas para as divisões das árvores de decisão.

Resumindo, quando o método de seleção RFE é utilizado em conjunto com as *Random Forests*, são calculadas as medidas de importância dos atributos que classificam os preditores do mais relevante para o menos. Os cálculos de importância podem ser baseados nos próprios critérios internos das *Random Forests* ou utilizando uma abordagem mais geral. Em cada estágio de busca, as variáveis preditoras menos importantes são eliminadas iterativamente antes da reconstrução do modelo. Uma vez que um novo modelo é criado, a função objetivo é estimada para aquele modelo, e o tamanho do subconjunto correspondente ao melhor valor da função objetivo é usado como modelo final (GUYON et al., 2002).

Capítulo 5

Metodologia

Neste capítulo são apresentadas as características do conjunto inicial de dados, as descrições da engenharia de atributos e do processo de rotulagem da variável resposta, além da divisão nos subconjuntos de treino e teste.

5.1 Caracterização do conjunto de dados inicial

O presente estudo se fundamenta em uma série temporal de dados financeiros, que reflete o comportamento do Ibovespa B3, o principal índice de ações da bolsa de valores brasileira. A base de dados compreende mais de 20 anos de observações, referente ao período de maio/1998 até junho/2021, e contém 5.715 registros obtidos a partir de registros da própria bolsa de valores. Cada instância do conjunto de dados representa um dia útil de negociação do índice e carrega informações sobre cinco variáveis: preço de abertura, preço de fechamento, preço máximo, preço mínimo e volume negociado. Não ocorreram dados ausentes. A Figura 25 apresenta o gráfico da evolução do Ibovespa B3, dentro do período analisado:

Figura 25 – Evolução do Ibovespa, principal índice de ações da B3



Fonte: Elaborado pelo autor (2022)

5.2 Atributos base – a primeira engenharia de atributos

A primeira transformação no conjunto de dados foi acrescentar os indicadores técnicos de mercado, apresentados no capítulo 2, e que utilizam em seus cálculos os preços de fechamento, máximo, mínimo e volume. Os indicadores foram os seguintes:

a) Rastreadores

- *Bollinger Bands – BB*
- *Average Directional Movement Index – ADX*
- *Parabolic SAR*

b) Osciladores

- *Moving Average Convergence / Divergence – MACD*
- *Commodity Channel Index – CCI*
- *Relative Strength Index – RSI*
- *Stochastic Oscillator – STO*

c) Volume

- *On-Balance Volume – OBV*
- *Accumulation / Distribution – A/D*
- *Volume Weighted Average Price - VWAP*

Em seguida, foi realizada a primeira engenharia de atributos com o propósito de criar novas variáveis preditoras. A premissa básica era capturar o cruzamento do preço de fechamento do índice com alguns dos indicadores técnicos. Por exemplo: se o parabólico SAR, estiver abaixo do preço de fechamento, a nova variável criada irá assumir valores positivos; caso esteja acima, os valores serão negativos. Com outros atributos, esta captura de importantes sinais para o modelo não necessita do preço de fechamento, ou seja, utiliza apenas o cruzamento entre linhas do próprio indicador, como, por exemplo, o movimento direcional ADX.

Vale observar que, no início da série, o índice valia cerca de 10.000 pontos; já no final, estava cotado em torno dos 130.000 pontos, ou seja, uma variação positiva de 500 pontos, no início, representava 5% de ganho no índice, ao passo que, essa mesma variação positiva de 500 pontos, em 2021, equivale, aproximadamente, 0,4% de ganho. A fim de expurgar esses efeitos ao longo da série temporal, a engenharia realizada com alguns dos indicadores exigiu a padronização dos valores obtidos, como no caso das bandas de bollinger – BB. Já os indicadores que assumem valores

dentro de uma faixa específica, como, por exemplo, o índice de força relativa – RSI e o estocástico – STO, que flutuam entre 0 e 100, essa padronização não foi necessária.

Ao concluir essa transformação inicial, foram excluídos os atributos preço de abertura, máximo, mínimo e volume negociado, permanecendo apenas o preço de fechamento junto com onze preditores que foram criados (a VWAP foi calculada para os períodos de 5 e 20 dias), totalizando assim 12 atributos base no conjunto de dados.

5.3 Atributos expandidos – a segunda engenharia de atributos

Em seguida, foi realizada a expansão dos 12 atributos base, através da implementação de três características específicas – Ruído, Gradiente e Posição relativa Máx-Mín, em janelas de curto e longo prazo, criando seis novos preditores para cada variável de entrada (ANGHINONI *et al.*, 2019). Essas características estão descritas a seguir, com a notação $s(d)$ indicando o novo atributo implementado:

- **Ruído:** os indicadores $N_S(d)$ e $N_L(d)$ correspondem aos ruídos de curto e longo prazo, respectivamente. Para cada ponto d da série temporal, o ruído é calculado como o desvio percentual da observação $s(d)$ em relação à média móvel de $s(d)$:

$$ruído_q(d) = \frac{s(d) - MA_q(d)}{MA_q(d)} = \frac{s(d)}{MA_q(d)} - 1$$

onde $MA_q(d)$ é a média móvel de $s(d)$ nos últimos q dias:

$$MA_q(d) = \frac{s(d) + s(d-1) + \dots + s(d-q+1)}{q}$$

Neste estudo, utilizamos $q_S = 10$ para cálculo do ruído de curto e $q_L = 50$ para ruído de longo prazo. Assim $N_S(d) = ruído_{10}(d)$ e $N_L(d) = ruído_{50}(d)$.

- **Gradiente:** os indicadores $G_S(d)$ e $G_L(d)$ correspondem aos gradientes de curto e longo prazo, respectivamente. Para cada ponto d da série, os gradientes de curto e longo prazo são calculados como a variação percentual da média móvel no dia d em relação à média móvel em $d-1$:

$$gradiente_q(d) = \frac{MA_q(d) - MA_q(d-1)}{MA_q(d-1)} = \frac{MA_q(d)}{MA_q(d-1)} - 1,$$

Assim, teremos $G_S(d) = gradiente_{10}(d)$ e $G_L(d) = gradiente_{50}(d)$.

- **Posição relativa Máx-Mín:** os indicadores $R_S(d)$ e $R_L(d)$ correspondem às posições máxima-mínima de curto e longo prazo, respectivamente. Para cada ponto d da série, a posição relativa máx-mín é calculada como a posição da observação $s(d)$ em relação à amplitude do valor na janela móvel:

$$relativa_q(d) = \frac{s(d) - \min(s[(d-q+1):d])}{\max(s[(d-q+1):d]) - \min(s[(d-q+1):d])},$$

Onde $s[(d-q+1):d]$ denota o subvetor $[s(d-q+1), s(d-q+2), \dots, s(d)]$. E, conforme anteriores, $R_S(d) = relativa_{10}(d)$ e $R_L(d) = relativa_{50}(d)$.

Neste ponto, o conjunto de dados passou a contar com 84 variáveis de entrada, ou seja, os 12 atributos base e mais seis atributos expandidos sobre cada um deles.

5.4 Rotulagem das classes na variável resposta

O próximo passo foi implementar a variável resposta para determinar, em cada observação do conjunto de dados, se naquele dia o mercado se encontrava em tendência de alta, de baixa ou sem tendência. Esse processo é conhecido como rotulagem da variável de saída. Um ponto importante a destacar é que, o papel do classificador a ser construído é prever, para cada dia t , a tendência futura do preço em um horizonte futuro $t+h$, com h denotando o tamanho do horizonte de predição. Desta forma, é fundamental que na construção do conjunto de treino e de teste ocorra o deslocamento da variável resposta neste horizonte h , para avaliar o comportamento futuro dos preços. O tamanho do horizonte h ficou definido, empiricamente, em cinco dias, após comparação dos erros estimados para diferentes janelas de deslocamento.

A estratégia utilizada para a rotulagem foi baseada no comportamento de três médias móveis – rápida, intermediária e lenta. As médias móveis apresentam bom desempenho em mercados com tendência, mas costumam gerar ruídos, ou falsos sinais, nos mercados sem tendência definida. Diante disso, a estratégia foi aplicada, simultaneamente, em dois tempos gráficos – diário e semanal. No tempo gráfico diário, o período utilizado para as médias móveis foram: rápida – 5 dias (m_5), intermediária – 15 dias (m_{15}) e lenta – 30 dias (m_{30}). Já no período semanal: média rápida – 15 dias (m_{15}), média intermediária – 30 dias (m_{30}) e média lenta – 60 dias (m_{60}).

A Figura 26 exemplifica uma típica evolução de três médias móveis – rápida, intermediária e lenta, sobre o preço de um ativo financeiro:

Figura 26 – Evolução de três médias móveis sobre o preço de um ativo



Fonte: Elaborado pelo autor (2022)

A estratégia consiste em observar a evolução das médias móveis, para determinar se o mercado se encontra em tendência de alta (classe +1), de baixa (classe -1) ou sem tendência (classe 0). Duas rotulagens preliminares, $YD(t)$ e $YS(t)$, são aplicadas sobre cada observação t da série, obedecendo os seguintes critérios.

a) Rotulagem considerando o tempo gráfico diário – notação: $YD(t)$

- Tendência de alta: se a média rápida (m_5) está acima da média intermediária (m_{15}) E acima da média lenta (m_{30}), então recebe classificação +1;
- Tendência de baixa: se a média rápida (m_5) está abaixo da média intermediária (m_{15}) E abaixo da média lenta (m_{30}), então recebe classificação -1;
- Sem tendência: em qualquer outra situação, recebe classificação 0.

b) Rotulagem considerando o tempo gráfico semanal – notação: $YS(t)$

- Tendência de alta: se a média rápida (m_{15}) está acima da média intermediária (m_{30}) E acima da média lenta (m_{60}), então recebe classificação +1;
- Tendência de baixa: se a média rápida (m_{15}) está abaixo da média intermediária (m_{30}) E abaixo da média lenta (m_{60}), então recebe classificação -1;
- Sem tendência: em qualquer outra situação, recebe classificação 0.

Diante desses critérios, a rotulagem final da variável resposta, $Y(t)$, considera em conjunto as rotulagens obtidas em cada tempo gráfico, respeitando as premissas apresentadas na Tabela 3. Observe que o tempo gráfico semanal $YS(t)$ predomina sobre o tempo gráfico diário $YD(t)$; já quando as rotulagens $YD(t)$ e $YS(t)$ assumem classes opostas, o rótulo final $Y(t)$ receberá classe 0, indicando mercado sem tendência.

Tabela 3 – Rotulagem da variável resposta do conjunto de dados

PREMISSAS PARA ROTULAGEM FINAL		
rotulagem diário $YD(t)$	rotulagem semanal $YS(t)$	rotulagem final $Y(t)$
+1	+1	+1 (alta)
0	+1	+1 (alta)
-1	+1	0 (lateral)
+1	0	0 (lateral)
0	0	0 (lateral)
-1	0	0 (lateral)
+1	-1	0 (lateral)
0	-1	-1 (baixa)
-1	-1	-1 (baixa)

Fonte: Elaborado pelo autor (2022)

Após o deslocamento da variável resposta e sua devida rotulagem, a função objetivo do algoritmo busca encontrar as relações existentes entre as variáveis de entrada que justifiquem o rótulo deslocado de cinco dias do atributo de saída, permitindo avaliar a capacidade preditiva das *Random Forests*. A representação formal deste processo de rotulagem e deslocamento h , está apresentada abaixo nas equações $YD(t)$, para o tempo gráfico diário, e $YS(t)$, para o tempo gráfico semanal:

$$YD(t) = \begin{cases} +1, & \text{se } m_5(t+h) > m_{15}(t+h) \wedge m_5(t+h) > m_{30}(t+h) \\ -1, & \text{se } m_5(t+h) < m_{15}(t+h) \wedge m_5(t+h) < m_{30}(t+h) \\ 0, & \text{caso contrário} \end{cases}$$

$$YS(t) = \begin{cases} +1, & \text{se } m_{15}(t+h) > m_{30}(t+h) \wedge m_{15}(t+h) > m_{60}(t+h) \\ -1, & \text{se } m_{15}(t+h) < m_{30}(t+h) \wedge m_{15}(t+h) < m_{60}(t+h) \\ 0, & \text{caso contrário} \end{cases}$$

Desta forma, o tratamento dos dados foi concluído após duas engenharias de atributos e com o deslocamento da variável resposta e sua rotulagem. O conjunto de dados teve uma discreta redução nas instâncias, devido às transformações, e passou a contar com 5.601 observações, compreendendo agora o período de 14/10/1998 a 11/06/2021, além das 84 variáveis de entrada e uma variável de saída.

5.5 Divisão do conjunto de dados e parametrização dos métodos

Nesta última etapa, foi feita a divisão da base de dados em conjunto de treino e conjunto de teste, para indução e validação dos algoritmos, respectivamente. A separação nos dois subconjuntos foi através da técnica *Holdout*, apresentada no capítulo 3, na proporção 75% treino e 25% teste. Vale ressaltar que, o *Holdout* implica numa amostragem aleatória das instâncias para a divisão; porém, no caso de séries temporais, os registros devem ser separados sequencialmente, ou seja, as primeiras 4.200 observações (75%) foram direcionadas para treino e as 1.401 observações finais (25%) foram destinadas para teste.

A Tabela 4 apresenta as proporções entre as três classes nos respectivos conjuntos de dados: inicial, treinamento e teste.

Tabela 4 – Proporções entre as classes nos conjuntos de dados

	Proporções entre as classes		
	classe -1	classe 0	classe +1
conjunto de dados inicial	27,57%	28,55%	43,88%
conjunto de treinamento	29,86%	27,67%	42,47%
conjunto de teste	20,70%	31,19%	48,11%

A principal parametrização dos algoritmos de *Random Forests*, tanto no núcleo do método VSURF quanto no núcleo do RFE, foi a quantidade de 2.000 árvores definida para cada modelagem, com cinco iterações em cada método. Também ficou estabelecido que a quantidade de atributos selecionada seria determinada pela mediana das iterações nos respectivos métodos. Os demais parâmetros, específicos do VSURF e do RFE, foram mantidos padrão.

As tarefas computacionais para seleção de atributos foram implementadas usando o ambiente estatístico R, versão 3.6.1. O computador utilizado para executar as rotinas possui as seguintes especificações: processador AMD Ryzen 1920x, de 3,5 GHz com 12 núcleos e 24 *threads*, placa-mãe Asrock x399, 64Gb de memória RAM DDR4 e sistema operacional Linux Ubuntu 18.04.5 LTS. Os dados e códigos utilizados no trabalho estão disponíveis em: https://github.com/smarteletto/mestrado_scx_usp.

Após os métodos VSURF e RFE selecionarem os atributos mais relevantes, cada um deles gerou um conjunto de dados específico, com a mesma quantidade de observações, mas diferenciado pela quantidade de variáveis preditoras selecionada. Em seguida, foi feita a validação nos respectivos conjuntos de teste, para obtenção dos resultados e avaliação de desempenho de cada modelo.

Capítulo 6

Resultados

Neste capítulo são demonstrados os resultados comparativos entre os métodos de seleção – VSURF e RFE, e também a análise das métricas de desempenho – tanto global quanto por classe. Além disso, são apresentadas a análise de estabilidade entre os principais atributos comuns que foram selecionados pelos métodos e a capacidade discriminante por classes desses atributos.

6.1 Comparação dos métodos de seleção e análise das métricas de desempenho

A partir do conjunto de dados com 84 atributos expandidos, o método VSURF selecionou a seguinte quantidade de atributos em cada uma das cinco iterações: 25, 25, 26, 26 e 26 atributos. O tempo médio de processamento foi de 1 hora e 15 minutos. A seleção final de 26 atributos foi determinada pela mediana das iterações, e a respectiva quantidade selecionada em cada etapa do método foi a seguinte:

- na limiarização: 82 variáveis selecionadas
- na interpretação: 57 variáveis selecionadas
- na predição: 26 variáveis selecionadas

Já a modelagem com o método RFE teve como solução ótima um conjunto de 64 atributos selecionados, também determinada pela mediana das cinco iterações. Cada uma apresentou o seguinte resultado: 61, 64, 64, 66 e 68 atributos. O tempo médio de processamento foi de 3 horas e 35 minutos.

Vale destacar a performance superior do método VSURF, tanto na redução de dimensionalidade, diminuindo de 84 atributos para 26 (69%) contra 64 do RFE (24%), quanto no tempo de processamento que foi 65% mais eficiente. A Tabela 5 apresenta estes resultados.

Tabela 5 – Número de atributos selecionados e tempos de processamento

	método VSURF		método RFE	
	nº atributos	tempo (minutos)	nº atributos	tempo (minutos)
iteração 1	25	71,7	61	215,2
iteração 2	25	77,4	64	213,4
iteração 3	26	75,4	64	213,4
iteração 4	26	73,6	66	217,7
iteração 5	26	78,4	68	219,0
resultados	26^(a)	75,3 / 2,7^(b)	64^(a)	215,7 / 2,5^(b)

(a) mediana dos atributos selecionados

(b) média / desvio-padrão dos tempos de processamento

Fonte: Elaborado pelo autor (2022)

Com a seleção de variáveis concluída pelos métodos VSURF e RFE, ficaram definidos os quatro conjuntos de dados a serem modelados pelas *Random Forests*. O objetivo da modelagem foi validar o algoritmo sobre as instâncias de teste de cada conjunto, para comparar as capacidades preditivas em diferentes dimensionalidades, utilizando métricas de desempenho global e por classe. Os quatro conjuntos de dados, com 5.601 observações, e diferentes atributos de entrada foram os seguintes:

- Conjunto 1 – dimensão de 12 atributos base;
- Conjunto 2 – dimensão de 84 atributos expandidos;
- Conjunto 3 – dimensão de 26 atributos selecionados pelo VSURF;
- Conjunto 4 – dimensão de 64 atributos selecionados pelo RFE.

As métricas de desempenho global de cada conjunto de dados, ordenados pela dimensionalidade, estão apresentadas na Tabela 6. Vale ressaltar que os limites inferior e superior indicados referem-se aos intervalos de 5% e 95%, respectivamente, obtidos durante a reamostragem via *bootstrap*, com 100 repetições, obtendo assim um intervalo de 90% de confiança, o que representa 1,64 desvios-padrão da média. Além disso, com exceção da base de dados de menor dimensionalidade, todos os conjuntos apresentaram sobreposição de seus intervalos, indicando que não existe diferença significativa entre as métricas de desempenho desses três modelos.

Tabela 6 – Métricas de desempenho global

MÉTRICAS de DESEMPENHO GLOBAL					
base12	Acurácia	MacroF1	PondF1	Matthews	Kappa
Média	0,7695	0,7537	0,7740	0,6403	0,6358
LimInf	0,7516	0,7334	0,7563	0,6128	0,6080
LimSup	0,7870	0,7725	0,7909	0,6672	0,6625
vsurf26	Acurácia**	MacroF1**	PondF1**	Matthews**	Kappa**
Média	0,8092	0,7957	0,8102	0,7001	0,6993
LimInf	0,7923	0,7784	0,7934	0,6744	0,6732
LimSup	0,8266	0,8143	0,8269	0,7273	0,7261
rfe64	Acurácia**	MacroF1**	PondF1**	Matthews**	Kappa**
Média	0,8219	0,8076	0,8227	0,7184	0,7180
LimInf	0,8059	0,7898	0,8064	0,6932	0,6927
LimSup	0,8394	0,8260	0,8399	0,7461	0,7459
expnd84	Acurácia**	MacroF1**	PondF1**	Matthews**	Kappa**
Média	0,8231	0,8095	0,8243	0,7207	0,7202
LimInf	0,8065	0,7913	0,8078	0,6947	0,6937
LimSup	0,8401	0,8271	0,8410	0,7462	0,7455

(***) sobreposição de limites com outros três conjuntos

(**) sobreposição de limites com outros dois conjuntos

(*) sobreposição de limites com apenas um conjunto

Fonte: Elaborado pelo autor (2022)

Analisando os resultados da Tabela 6, é possível verificar que o conjunto de dados de menor dimensionalidade, com 12 indicadores base, apresentou as métricas de menor desempenho, embora seus valores de acurácia e pontuações F1 tenham ficado acima de 75% e os coeficientes de Matthews e Kappa acima de 60%. O conjunto de dados expandido para 84 atributos foi o que apresentou a melhor performance, visto que a maior quantidade de atributos de entrada permitiu ao modelo capturar de forma mais eficiente as relações entre os preditores e a variável resposta.

Quando analisamos as métricas dos conjuntos de dados obtidos a partir dos métodos de seleção, percebemos que o desempenho do RFE com 64 atributos foi similar ao do conjunto expandido com 84 atributos, e levemente superior ao VSURF com 26 atributos. Entretanto, a pequena vantagem do RFE sobre o VSURF, em relação às médias das variáveis preditoras, não parece compensar a grande

superioridade do VSURF em termos de redução de dimensionalidade e de performance no tempo de processamento.

Com relação às métricas para avaliar o desempenho por classe foram utilizadas a precisão e o *recall* para comparação entre tendência de baixa (classe -1), sem tendência (classe 0) e tendência de alta (classe 1). Vale ressaltar que a medida de precisão representa a proporção de positivos previstos que são, realmente, positivos. Desta forma, um alto valor de precisão significa a chance de ganho se um investidor abrir uma operação de compra ou venda. Já a medida de *recall*, representa a proporção de positivos reais que foram classificados corretamente. Assim, um alto valor de *recall*, significa o percentual de oportunidades de compra ou venda que seriam aproveitadas pelo investidor. Diante disso, é importante que os classificadores obtenham maior precisão e *recall* nas classes -1 e +1, que são as que determinariam a abertura ou fechamento de operações. Os resultados obtidos a partir das métricas de desempenho por classe estão reunidos na Tabela 7.

Tabela 7 – Métricas de desempenho por classe

MÉTRICAS de DESEMPENHO por CLASSE						
base12	Prec-1***	Prec0	Prec1***	Recall-1	Recall0***	Recall1*
Média	0,7972	0,6129	0,8982	0,6643	0,7598	0,8221
LimInf	0,7553	0,5809	0,8787	0,6193	0,7259	0,7980
LimSup	0,8390	0,6477	0,9178	0,7071	0,7931	0,8457
vsurf26	Prec-1***	Prec0**	Prec1***	Recall-1**	Recall0***	Recall1***
Média	0,7642	0,6967	0,9089	0,8485	0,6994	0,8634
LimInf	0,7270	0,6597	0,8903	0,8129	0,6642	0,8416
LimSup	0,8019	0,7321	0,9277	0,8808	0,7347	0,8839
rfe64	Prec-1***	Prec0**	Prec1***	Recall-1**	Recall0***	Recall1**
Média	0,7869	0,7160	0,9097	0,8307	0,7228	0,8823
LimInf	0,7484	0,6803	0,8913	0,7939	0,6872	0,8629
LimSup	0,8243	0,7528	0,9282	0,8658	0,7584	0,9025
expnd84	Prec-1***	Prec0**	Prec1***	Recall-1**	Recall0***	Recall1**
Média	0,7915	0,7135	0,9141	0,8280	0,7349	0,8782
LimInf	0,7516	0,6797	0,8960	0,7902	0,7000	0,8566
LimSup	0,8300	0,7476	0,9314	0,8616	0,7704	0,8991

(***) sobreposição de limites com outros três conjuntos

(**) sobreposição de limites com outros dois conjuntos

(*) sobreposição de limites com apenas um conjunto

Da Tabela 7, vemos que a classe -1 – tendência de baixa, apresentou uma métrica de precisão em torno de 80%, ou seja, praticamente a mesma em todos os modelos, já que ocorreu sobreposição entre todos intervalos, indicando que não houve diferença significativa entre os conjuntos de dados. Na medida de *recall* desta classe, houve um menor desempenho no conjunto de menor dimensionalidade – 12 atributos, com valor em torno de 66%, enquanto nos demais conjuntos a métrica ficou na ordem de 83%, sem diferença significativa entre eles.

Para a classe 0 – sem tendência, a análise vertical da tabela 7 mostra que a métrica de precisão foi menor no conjunto de menor dimensionalidade – 12 atributos, com valor em torno de 61%, enquanto nos demais modelos assumiu valores da ordem de 70%, sem diferença significativa entre eles, devido à sobreposição dos intervalos. A medida de *recall* desta classe também não apresentou diferença estatística entre os modelos, apresentando valores em torno de 72%.

Na classe 1 – tendência de alta, os resultados da métrica precisão foram da ordem de 90%, com os valores de todos os modelos não apresentando diferença estatisticamente significativa entre si, devido à sobreposição entre seus limites inferior e/ou superior. E a medida de *recall* desta classe, teve um resultado de 82% no modelo com 12 atributos, sem diferença significativa com o modelo VSURF de 26 atributos, enquanto os demais modelos a métrica assumiu valores em torno de 88%, sem diferenças entre eles.

Já numa análise horizontal da Tabela 7, podemos observar que a métrica de precisão da classe 1 – tendência de alta, foi a que apresentou melhor desempenho em todos os conjuntos de dados, apresentando valores da ordem de 90%. E com relação à medida de *recall*, também numa análise horizontal, foi a mesma classe 1 que obteve os melhores desempenhos, assumindo valores da ordem de 85%.

6.2 Estabilidade e capacidade discriminante dos atributos selecionados

Antes de apresentar os resultados da análise de estabilidade, serão relacionados na Tabela 8 os atributos mais importantes que foram selecionados pelos métodos, com suas respectivas siglas. Basicamente, cada sigla informa qual indicador técnico foi utilizado na criação daquele atributo. Além disso, alguns deles estão acompanhados de um sufixo, indicando que foram obtidos da engenharia de expansão, ou seja, a partir das três características implementadas: ruído de curto prazo (Ns), ruído de longo prazo (NI), gradiente de curto prazo (Gs), gradiente de longo prazo (GI), posição relativa de curto prazo (Rs) e posição relativa de longo prazo (RI).

Tabela 8 – Identificação dos principais atributos selecionados

ATRIBUTO	INDICADOR TÉCNICO	SUFIXO
PriceCl.NI	preço de fechamento	ruído longo
PriceCl.RI	preço de fechamento	posição relativa longo
RSI	índice de força relativa	-
ClxVWAP20	preço médio ponderado pelo volume	-
PriceCl.GI	preço de fechamento	gradiente longo
ClxBB	bandas de bollinger	-
ClxVWAP20.Ns	preço médio ponderado pelo volume	ruído curto
RSI.NI	índice de força relativa	ruído longo
RSI.RI	índice de força relativa	posição relativa longo
pnADX.RI	movimento direcional	posição relativa longo
pnADX	movimento direcional	-
ClxVWAP20.RI	preço médio ponderado pelo volume	posição relativa longo
ClxVWAP20.NI	preço médio ponderado pelo volume	ruído longo
RSI.Ns	índice de força relativa	ruído curto
MACD	movimento convergente-divergente	-
CCI	índice do canal de commodities	-
ClxBB.Ns	bandas de bollinger	ruído curto
CCI.Ns	índice do canal de commodities	ruído curto
RSI.GI	índice de força relativa	gradiente longo
pnADX.Gs	movimento direcional	gradiente curto
PriceCl.Ns	preço de fechamento	ruído curto
ClxBB.RI	bandas de bollinger	posição relativa longo
pnADX.Ns	movimento direcional	ruído curto
AD.NI	acumulação-distribuição	ruído longo
ClxSAR.Rs	parabólico sar	posição relativa curto

Fonte: Elaborado pelo autor (2022)

A análise de estabilidade de atributos foi realizada entre os dois métodos, com o objetivo de comparar as variáveis mais importantes de cada grupo selecionado. Para isso, foi verificado se os vinte atributos mais bem classificados pelo método VSURF, também estavam presentes nas primeiras vinte posições do método RFE. Em seguida, os atributos receberam uma pontuação, de acordo com a posição que ocupavam em cada método, o que permitiu a inferência de um ranking de importância geral para os atributos selecionados.

Esta pontuação conferida a cada atributo equivale à soma das respectivas posições ocupadas pelo mesmo em cada método de seleção, ou seja, quanto menor a pontuação, melhor é o posicionamento (classificação ascendente); por exemplo: o atributo PriceCl.NI, que foi o melhor classificado com pontuação 3, ocupava a primeira posição de importância no método VSURF e a segunda posição no método RFE e, assim, sucessivamente. Se um atributo não aparece num dos métodos de seleção, o mesmo recebe 100 pontos de penalização pela ausência, de modo que este critério possa justificar as pontuações mais baixas dos atributos comuns.

Na Tabela 9 estão os resultados obtidos dessa análise de estabilidade, sendo que as colunas desta tabela estão descritas abaixo:

- Rank: classificação ascendente dos 20 melhores atributos selecionados em cada método (hachurado em verde, os 15 atributos comuns);
- Atributo: identificação de cada atributo selecionado (hachurado em amarelo, os 15 atributos comuns; hachurado em vermelho, os atributos ausentes no VSURF, mas presentes no RFE; hachurado em azul, os indicadores presentes no VSURF, mas ausentes no RFE);
- VSURF: posição do atributo, por ordem de importância, no VSURF;
- RFE: posição do atributo, por ordem de importância, no RFE;
- Pontuação: *score* referente à soma das respectivas posições em cada um dos métodos;
- Tipo: classificação dos 15 atributos comuns por tipo: R – rastreadores, O – osciladores e V – volume.

Tabela 9 – Estabilidade dos atributos entre os métodos de seleção

RANK	ATRIBUTO	VSURF	RFE	PONTUAÇÃO	TIPO
1	PriceCl.NI	1	2	3	R
2	PriceCl.RI	2	1	3	R
3	RSI	3	3	6	O
4	ClxVWAP20	4	5	9	V
5	PriceCl.GI	8	4	12	R
6	ClxBBB	5	9	14	R
7	ClxVWAP20.Ns	10	6	16	V
8	RSI.NI	7	11	18	O
9	RSI.RI	6	13	19	O
10	pnADX.RI	12	14	26	R
11	pnADX	9	18	27	R
12	ClxVWAP20.RI	13	17	30	V
13	ClxVWAP20.NI	17	16	33	V
14	RSI.Ns	18	19	37	O
15	MACD	20	20	40	O
16	CCI	100	7	107	
17	ClxBBB.Ns	100	8	108	
18	CCI.Ns	100	10	110	
19	RSI.GI	11	100	111	
20	pnADX.Gs	100	12	112	
21	PriceCl.Ns	14	100	114	
22	ClxBBB.RI	15	100	115	
23	pnADX.Ns	100	15	115	
24	AD.NI	16	100	116	
25	ClxSAR.Rs	19	100	119	

Fonte: Elaborado pelo autor (2022)

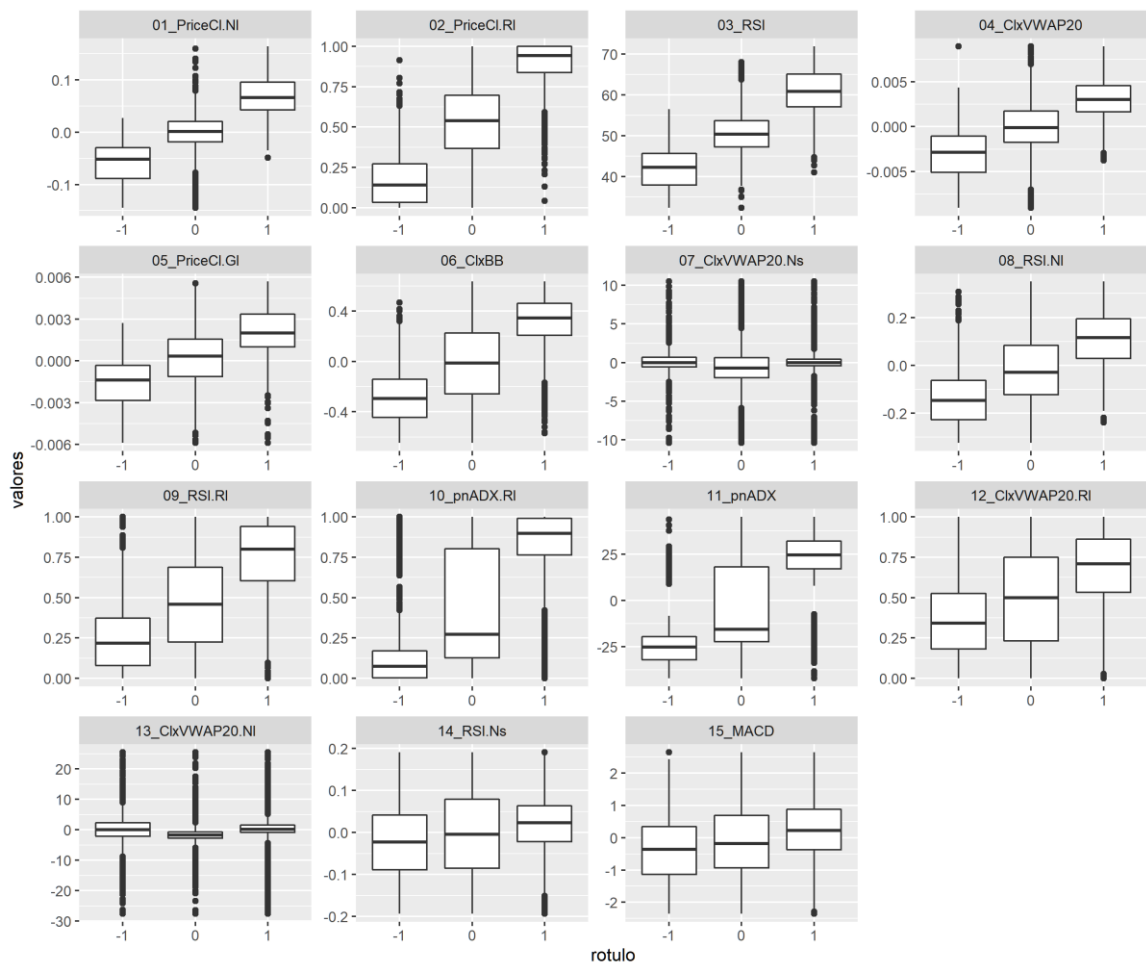
A partir da Tabela 9, foram feitas as seguintes considerações sobre os resultados obtidos da análise de estabilidade:

- a) dos vinte atributos mais importantes de cada grupo selecionado, quinze deles foram comuns aos dois métodos, indicando uma estabilidade de 75% entre tais atributos. Apenas dois atributos ocuparam a mesma posição de importância, tanto no VSURF quanto no RFE – o atributo RSI, na posição 3, e o MACD na posição 15;

- b) sobre os quinze atributos comuns, selecionados pelos dois métodos, podemos observar que seis deles são indicadores do tipo rastreador (R), cinco são do tipo oscilador (O) e quatro são indicadores de volume (V);
- c) ainda sobre os atributos comuns, cinco deles são oriundos dos atributos base, representando 42% de um total de 12, enquanto dez atributos foram originados a partir da expansão realizada sobre o conjunto de dados, representando 14% de um total de 72.

Para finalizar, também foi realizada uma análise da capacidade discriminante, nas respectivas classes, de cada atributo comum aos dois métodos de seleção. Inicialmente, para cada atributo comum, foi construído um grupo com três *boxplots*, os quais representam cada uma das classes, indicados no eixo x. Já os valores que cada atributo comum assume no conjunto de dados estão representados no eixo y. A Figura 27 ilustra a capacidade discriminante por classe de cada atributo comum:

Figura 27 – Capacidade discriminante por classe de cada atributo comum



A intuição por trás desta análise é identificar quais atributos selecionados foram capazes de discriminar, de maneira efetiva, as tendências de alta (classe +1), as de baixa (classe -1) e os mercados sem tendência (classe 0). Numa análise visual de cada grupo de *boxplots*, é possível perceber que as melhores separações foram obtidas pelos três primeiros atributos. Tomando como exemplo o melhor atributo selecionado – PriceCI.NI, percebe-se que o terceiro quartil do *boxplot* da classe -1 não se sobrepõe ao primeiro quartil do *boxplot* da classe 0; além disso, o terceiro quartil do *boxplot* da classe 0 não se sobrepõe ao primeiro quartil do *boxplot* da classe +1. Esta situação também é observada, tanto no segundo quanto no terceiro melhor atributo selecionado, respectivamente – PriceCI.RI e RSI. Embora nenhum dos demais atributos tenham se comportado de maneira semelhante, alguns deles tiveram resultados discriminantes bastante satisfatórios.

Capítulo 7

Conclusões

Este trabalho apresentou uma comparação empírica entre os métodos de seleção de atributos VSURF (*Variable Selection Using Random Forests*) e RFE (*Recursive Feature Selection*), no contexto da previsão de tendências de uma série temporal financeira com aproximadamente 25 anos de observações, representada pelo principal índice de ações da B3 – Ibovespa.

Tanto o método de seleção VSURF quanto o RFE, foram capazes de reduzir o número de atributos de entrada do conjunto de dados, sendo que o VSURF realizou o processo em tempo médio de 1h15m contra 3h35m do RFE – uma superioridade de desempenho computacional de quase 3 vezes. A redução efetiva do número de atributos do VSURF também foi superior, da ordem de 69%, passando de 84 atributos do conjunto de dados para 26 atributos. Já no método RFE, essa redução foi menos efetiva, da ordem de 24%, passando de 84 atributos para 64 atributos.

Além disso, nenhum dos métodos causou perda na capacidade preditiva do modelo quando comparados aos conjuntos de controle – com 12 e 84 variáveis. Embora os valores médios das métricas de desempenho obtidas no método RFE tenham sido ligeiramente superiores às obtidas pelo VSURF (como a acurácia global que registrou 82% de acertos no RFE, contra 80% no VSURF, e o índice Kappa que apresentou um valor de 72% no RFE, contra 69% no VSURF), nenhuma das diferenças observadas foi significativa do ponto de vista estatístico. Assim, no estudo de caso analisado, essa leve superioridade nas métricas do RFE não chega a compensar a velocidade de processamento e a efetiva redução de dimensionalidade do VSURF. Este é um achado interessante e demonstra que o método VSURF, proposto na literatura mais recentemente do que o tradicional método RFE, tende a se popularizar nos próximos anos em aplicações utilizando *Random Forests*.

Com relação à estabilidade dos atributos mais importantes selecionados pelas técnicas, foi possível verificar equilíbrio e consistência entre os mesmos, ou seja, os atributos comuns aos métodos ocuparam as 15 primeiras posições em ambos subconjuntos, além do fato dessas posições estarem bastante próximas no ranking de classificação. Já a capacidade discriminante dos atributos selecionados se mostrou superior entre aqueles que ocuparam as primeiras posições no ranking de classificação.

A rotulagem da variável resposta, através do cruzamento de médias móveis, foi a estratégia adotada por viabilizar a combinação de tempos gráficos diferentes – diário e semanal, mas existem alternativas que podem ser exploradas para este processo. Futuras pesquisas serão desenvolvidas com outros ativos financeiros, utilizando tempos gráficos intradiários e avaliando novos algoritmos de modelagem. Assim será possível identificar potenciais ameaças às conclusões propostas por este trabalho, bem como responder eventuais questões que não tenham sido consideradas.

Referências

- AHMED, N. K., ATIYA, A. F., GAYAR, N. E., EI-SHISHINY, H. **An empirical comparison of machine learning models for time series forecasting.** *Econometric Reviews*, vol. 29, p. 594-621. 2010
- ANGHINONI, L., ZHAO, L., DONGHONG, J., PAN, H. **Times series trend detection and forecasting using complex network topology analysis.** *Elsevier Neural Networks*, vol. 117, p. 295-306. 2019
- ARÉVALO, R., GARCIA, J., GUIJARRO, F., PERIS, A. **A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting.** *Expert Systems with Applications*, vol. 81, p. 177-192. 2017
- BAR-YAM, Y. **Dynamics of Complex Systems: studies in nonlinearity.** *Westview Press*. Colorado. 2003
- BERRAR, D. **Cross-validation.** *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, vol. 1, p. 542-545. 2018
- BLUM, A. L., LANGLEY, P. **Selection of relevant features and examples in machine learning.** *Elsevier Artificial Intelligence*, p. 245-271. 1997
- BREIMAN, L. **Random Forests.** *Machine Learning Kluwer Academics*, vol. 45, p. 5-32. 2001
- BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., STONE, C. J. **Classification and regression trees.** *Chapman & Hall*. 1998
- BROOKS, A. **Trading Price Action – Trends.** *John Wiley & Sons*. New Jersey. 2012
- BROWNE, M. W. **Cross-validation methods.** *Journal of Mathematical Psychology*, vol. 44, p. 108-132. 2000
- COLBY, R. W. **The Encyclopedia of Technical Market Indicators.** *McGraw-Hill*. New York. 2010
- CUNNINGHAM, J. P., GHAHRAMANI, Z. **Linear dimensionality reduction: survey, insights, and generalizations.** *Journal of Machine Learning Research*, vol. 16. 2015

- DONOHO, D. L. **High-dimensional data analysis: the curses and blessings of dimensionality**. *Math challenges of 21st century*. University of Stanford. 2000
- EATWELL, J., MILGATE, M., NEWMAN, P. **Finance**. *The Macmillian Press Limited*, p. 127-134. England. 1994
- FACELI, K., LORENA, A. C., GAMA, J., de CARVALHO, A. C. P. L. F. **Inteligência Artificial – uma abordagem de aprendizado de máquina**. LTC. Rio de Janeiro. 2011
- FAN, J., LI, R. **Statistical challenges with high-dimensionality: feature selection in knowledge discovery**. *arXiv Math*. 2006
- FONTI, V., BELITSER, E. **Feature selection using LASSO**. *University of Amsterdam*. 2017
- FRIZZARINI, C. **Algoritmo para indução de árvores de classificação para dados desbalanceados**. *Dissertação, Universidade de São Paulo*. 2013
- GENUER, R., POGGI, J. M. **Random Forests with R**. Springer. 2020
- GRANDINI, M., BAGLI, E., VISANI, G. **Metrics for multi-class classification: an overview**. *University of Bologna*. 2020
- GUYON, I. WESTON, J., BARNHILL, S., VAPNIK, V. **Gene selection for cancer classification using support vector machine**. *Machine Learning*, vol. 46. p. 389-422. 2002
- KIRKPATRICK II, C. D., DAHLQUIST, J. **Technical Analysis**. *Pearson Education*. New Jersey. 2010
- KIRKPATRICK II, C. D., DAHLQUIST, J. A. **Technical analysis: the complete resource for financial market technicians**. *Editora FT Press*. 2010
- KUHN, M., JOHNSON, K. **Applied Predictive Modeling**. *Springer*. 2016
- KUHN, M., JOHNSON, K. **Feature engineering and selection – a practical approach for predictive models**. *CRC Press*. 2020
- LAURETTO, M. S. **Árvores de classificação para escolha de estratégias de operação em mercados de capitais**. *Dissertação, Universidade de São Paulo*. 1996

LEE, J. A., VERLEYSEN, M. **Nonlinear dimensionality reduction techniques: a data structure preservation approach.** *Springer.* 2021

LIU, H., MOTODA, H. **Computational methods of feature selection.** *Chapman & Hall.* New York. 2008

LO, A. W., MACKINLAY, A. C. **Stock market prices do not follow random walks: evidence from a simple specification test.** *National Bureau of Economic Research.* Cambridge. 1987

MANOJLOVIC, T., STAJDUHAR, I. **Predicting stock market trends using Random Forests: a sample of the Zagreb stock exchange.** *IEEE 38th International Convention on Information and Communication Technology,* p. 1189-1193. 2015

MITCHELL, M. **Artificial Intelligence – a guide for thinking humans.** *Farrar, Straus and Giroux.* New York. 2019

MURPHY, J. J. **Technical Analysis of the Financial Markets.** *New York Institute of Finance.* New York. 1999

NGUYEN, T. H., SHIRAI, K., VELCIN, J. **Sentiment Analysis on social media for stock movement prediction.** *Expert Systems with Applications,* vol. 42(24), p. 9603-9611. 2015

PATEL, J., SHAH, S., THAKKAR, P., KOTECHEA, K. **Predicting stock market index using fusion of machine learning techniques.** *Expert Systems with Applications,* vol. 42(4), p. 2162-2172. 2015

PEARL, J. **Probabilistic Reasoning in Intelligent Systems.** *Morgan Kaufmann Publications Inc.* California. 1988

PHALAK, P., BHANDARI, K., SHARMA, R. **Analysis of Decision Tree – a survey.** *International Journal of Engineering Research & Technology,* vol. 3. 2014

ROKACH, L., MAIMON, O. **Top-down induction of decision trees classifiers – a survey.** *IEEE Transactions on Systems and Cybernetics,* vol. 35. 2005

RUSSELL, S., NORVIG, P. **Artificial Intelligence: a modern approach.** *Pearson Education Limited.* England. 2016

SARKER, I. H. **Machine Learning: algorithms, real-world applications and research directions.** *SN Computer Science*. 2021

SAYAMA, H. **Introduction to Modeling and Analysis of Complex Systems.** *Binghamton University*. New York. 2015

SVETNIK, V., LIAW, A., TONG, C., CULBERSON, C., SHERIDAN, R., FEUSTON, B. **Random Forest: a classification and regression tool for compound classification and QSAR modeling.** *Journal of Chemical Information and Computer Science*, vol. 43, p. 1947-1958. 2003

TAEHO, J. **Machine Learning Foundations: supervised, unsupervised, and advanced learning.** *Springer*. 2021

TAYLOR, M. C. **The moment of Complexity: emerging network culture.** *University of Chicago*. Chicago. 2003

TAYLOR, S. **Modeling Financial Time Series.** *John Wiley & Sons*. England. 1986

TSAY, R. S. **Analysis of Financial Time Series.** *John Wiley & Sons*. England. 2002

TYRALIS, H., PAPACHARALAMPOUS, G. **Variable selection in time series forecasting using random forests.** *Algorithms, MDPI*. 2017

WEBB, A. R., COPSEY, K. D. **Statistical pattern recognition.** *John Wiley & Sons*. England. 2011

WENG, B., LU, L., WANG, X., MEGAHED, F. M., MARTINEZ, W. **Predicting short-term stock prices using ensemble methods and online data sources.** *Expert Systems and Applications*, vol. 112, p. 258-273. 2020

WITTEN, I., FRANK, E., HALL, M. A., PAL, C. J. **Data Mining: practical machine learning tools and techniques.** *Elsevier*. 2017

YAOHAO, P. **Big Data, machine learning and challenges of high dimensionality in financial administration.** *Tese, Universidade de Brasília*. 2019

ZHENG, T., ZIQIN, Y., GUANGWEI, Z. **Stock selection with random forest: an exploitation of excess return in the Chinese stock market.** *Heliyon, Elsevier*. 2019