



UNIVERSIDADE FEDERAL DE PERNAMBUCO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Rogério do Amaral Armstrong

Análise do desempenho de diferentes modelos de Machine Learning na previsão de preços de ativos de renda variável

RECIFE
2024

Rogério do Amaral Armstrong

**Análise do desempenho de diferentes modelos de Machine Learning
na previsão de preços de ativos de renda variável**

Trabalho de Conclusão de Curso apresentado ao Curso Engenharia de Produção da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Engenheiro de Produção.

Orientador: Prof. Dr. Raphael Harry Frederico Ribeiro Kramer

RECIFE
2024

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Armstrong, Rogerio do Amaral.

Análise do desempenho de diferentes modelos de Machine Learning na
previsão de preços de ativos de renda variável / Rogerio do Amaral Armstrong. -
Recife, 2024.

49 p.

Orientador(a): Raphael Harry Frederico Ribeiro Kramer
(Graduação) - Universidade Federal de Pernambuco, Centro de Tecnologia e
Geociências, , 2024.

1. Mercado Financeiro. 2. Machine Learning. 3. Previsão de Preços. 4.
Indicadores Fundamentalistas. 5. Indicadores Técnicos. I. Kramer, Raphael
Harry Frederico Ribeiro. (Orientação). II. Título.

620 CDD (22.ed.)

Rogério do Amaral Armstrong

**Análise do desempenho de diferentes modelos de Machine Learning
na previsão de preços de ativos de renda variável**

Trabalho de Conclusão de Curso apresentado ao Curso Engenharia de Produção da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Engenheiro de Produção.

Aprovado em: 18/03/2024

BANCA EXAMINADORA

Prof. Dr. Raphael Harry Frederico Ribeiro Kramer (Orientador)

Universidade Federal de Pernambuco

Profa. Dra. Maísa Mendonça Silva (Examinador Interno)

Universidade Federal de Pernambuco

Profa. Dra. Isis Didier Lins (Examinador Interno)

Universidade Federal de Pernambuco

Resumo

A utilização de diferentes técnicas para tentar prever os preços de ativos de renda variável tem sido um tópico de extrema relevância no cenário financeiro, especialmente na era da tecnologia e das transformações digitais que vem gerando uma massa cada vez maior de dados acerca desses ativos. Nesse contexto, o uso de modelos de *Machine Learning* para realizar tais previsões ganhou destaque como uma ferramenta promissora para investidores e analistas, dada a capacidade dos modelos de identificar padrões complexos nos dados e adaptar-se a novas informações, tornando-os uma alternativa atrativa e que pode ser amplamente explorada nesse mercado. A volatilidade dos mercados e a influência de diversos fatores econômicos, políticos e sociais tornam o comportamento dos preços altamente imprevisível. Diante dessa incerteza, o desenvolvimento de métodos precisos e eficazes de previsão se tornou um objetivo fundamental para a tomada de decisões informadas e a otimização de estratégias de investimento. Este trabalho investiga o desempenho de modelos de *Machine Learning* na previsão de preços de ativos de renda variável. Ao analisar a precisão e eficácia desses modelos, o estudo oferece *insights* valiosos sobre a aplicabilidade de técnicas de *Machine Learning* no campo das finanças, destacando o potencial dessas ferramentas na análise de mercado e tomada de decisão em investimentos.

Palavras-chave: Mercado Financeiro; *Machine Learning*; Previsão de Preços; Indicadores Fundamentalistas; Indicadores Técnicos

Lista de Ilustrações

Figura 1 – Composição do Sistema Financeiro Nacional	11
Figura 2 - Categorias das variáveis utilizadas para previsão de preços de ações (adaptado de Kumbure et al, 2022)	16
Figura 3 – Logomarca Petrobras	20
Figura 4 – Logomarca Amazon	21
Figura 5 - Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo <i>SVR</i> e apenas indicadores fundamentalistas	33
Figura 6 – Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo <i>SVR</i> , indicadores técnicos e indicadores fundamentalistas	34
Figura 7 – Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo <i>Random Forest</i> e apenas indicadores técnicos	35
Figura 8 - Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo <i>Random Forest</i> e apenas indicadores fundamentalistas	36
Figura 9 - Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo <i>Random Forest</i> , indicadores técnicos e indicadores fundamentalistas	36
Figura 10 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo <i>SVR</i> e apenas indicadores técnicos	37
Figura 11 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo <i>SVR</i> e apenas indicadores fundamentalistas	38
Figura 12 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo <i>SVR</i> , indicadores técnicos e indicadores fundamentalistas	39
Figura 13 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo <i>Random Forest</i> e apenas indicadores técnicos	40
Figura 14 – Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo <i>Random Forest</i> e apenas indicadores fundamentalistas	40
Figura 15 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo <i>Random Forest</i> , indicadores técnicos e indicadores fundamentalistas	41

Lista de Tabelas

Tabela 1 – Indicadores selecionados para utilização nos modelos de <i>Machine Learning</i>	27
Tabela 2 – Resultado da análise comparativa	42

Lista de Abreviaturas Siglas

AWS	Amazon Web Services
BACEN	Banco Central do Brasil
LAJIDA	Lucro Antes dos Juros, Impostos, Depreciação e Amortização
MME	Média Móvel Exponencial
MMS	Média Móvel Simples
<i>OBV</i>	<i>On-Balance Volume</i>
<i>RBF</i>	<i>Radial Basis Function</i>
<i>ROE</i>	<i>Return On Equity</i>
<i>RSI</i>	<i>Relative Strength Index</i>
SFN	Sistema Financeiro Nacional
SUSEP	Superintendência de Seguros Privados
<i>SVR</i>	<i>Support Vector Regression</i>
<i>SVM</i>	<i>Support Vector Machine</i>

SUMÁRIO

1. INTRODUÇÃO	6
1.1. Justificativa e relevância	6
1.2. Objetivos	7
1.3. Metodologia	7
1.4. Estrutura do trabalho	8
2. REFERENCIAL TEÓRICO	10
2.1. Composição do mercado financeiro brasileiro.....	10
2.2. Análise de ativos de renda variável.....	12
2.3. Técnicas de Machine Learning	13
2.3.1. <i>Support Vector Regression (SVR)</i>	14
2.3.2. <i>Random Forest</i>	14
3. REVISÃO DA LITERATURA	16
4. DESENVOLVIMENTO DOS MODELOS	19
4.1. Empresas analisadas.....	19
4.2. Tipos de indicadores	22
4.3. Indicadores fundamentalistas	22
4.4. Indicadores técnicos	25
4.5. <i>Support Vector Regression (SVR)</i>	27
4.6. <i>Random Forest</i>	28
5. ANÁLISE COMPARATIVA	30
6. CONCLUSÃO	43
REFERÊNCIAS	44

1. INTRODUÇÃO

Nos últimos anos, é observado uma popularização do mercado financeiro no Brasil devido ao surgimento de influenciadores digitais no nicho de finanças e investimentos, maior acessibilidade às informações financeiras, avanços tecnológicos como plataformas de negociação online e aplicativos de investimento. Além do período recente com taxas de juros mais baixas, que incentivaram a busca por alternativas de investimento com maior rentabilidade.

Todos esses fatores foram responsáveis por democratizar o acesso a diferentes produtos financeiros, atraindo um novo perfil de investidores e diversificando o portfólio de investimentos no país.

A busca por métodos para prever os valores de ativos de renda variável é crucial no ambiente financeiro atual, especialmente com o avanço tecnológico e a digitalização, que produzem grandes volumes de dados sobre esses ativos. Essas estratégias visam aperfeiçoar a distribuição de recursos, buscando maximizar os retornos e minimizar os riscos.

Nesse cenário, as técnicas de *Machine Learning* têm se destacado como recursos valiosos para investidores e analistas, devido à sua habilidade em reconhecer padrões complexos nos dados e se ajustar a novas informações, apresentando-se como uma opção eficiente e aplicável à realidade do mercado financeiro.

A imprevisibilidade dos preços, influenciada por variados fatores econômicos, políticos e sociais, eleva a complexidade do mercado. Portanto, criar métodos de previsão precisos e eficientes é essencial para embasar decisões e otimizar estratégias de investimento sob tais condições de incerteza.

Este trabalho tem como propósito explorar e comparar a utilização de diferentes modelos de *Machine Learning* fazendo uso de diferentes *inputs* (e.g., baseados em indicadores técnicos e fundamentalistas), na previsão de preços de ativos de renda variável,

1.1. Justificativa e relevância

Ao longo dos anos, diversas abordagens têm sido empregadas na tentativa realizar previsões assertivas acerca dos preços de ativos de renda variável. Inicialmente, métodos tradicionais como a análise técnica e fundamentalista eram predominantes, porém, com o advento da era digital e o acesso a grandes volumes de dados, surgiu a oportunidade de aplicar técnicas de *Machine Learning* para aprimorar as previsões financeiras.

As motivações que levaram a escolha deste tema são diversas e fundamentadas. Em primeiro lugar, a crescente importância do mercado financeiro no cenário econômico global reforça a relevância de estudos que possam contribuir para aprimorar as técnicas de previsão de preços. Além disso, a utilização de modelos de *Machine Learning* para essa finalidade tem demonstrado resultados promissores em diversas áreas, incrementando o interesse em investigar a aplicabilidade desses modelos em um contexto dinâmico e desafiador como o mercado de renda variável.

Além disso, há uma busca por soluções inovadoras e eficazes para os desafios presentes no mercado, sendo fatores que também impulsionam a realização deste trabalho. A utilização de indicadores técnicos e fundamentalistas como inputs para diferentes modelos de *Machine Learning* pode fornecer *insights* valiosos, abrindo novas perspectivas para aprimorar a capacidade de previsão e, conseqüentemente, a tomada de decisões estratégicas em investimentos.

1.2. Objetivos

Esse trabalho tem por objetivo central analisar os *outputs* obtidos pela aplicação de diferentes modelos de *Machine Learning* implementados, combinados aos diferentes *inputs*, culminando em uma dissertação e avaliação acerca da performance de cada um deles na previsão de preços de ativos de renda variável.

São propostos os seguintes objetivos específicos:

- Revisar a literatura acerca dos modelos de *Machine Learning* utilizados para previsão de preços de renda variável;
- Implementar algoritmos utilizando linguagem de programação para importação de dados e cálculo de indicadores;
- Implementar os algoritmos de *Machine Learning* selecionados para previsão de preços dos ativos de renda variável;
- Sintetizar os resultados obtidos pelos modelos implementados em forma de gráficos e/ou tabelas para elaboração de análise comparativa.

1.3. Metodologia

De acordo com Gil (2008), o método dedutivo parte de princípios incontestáveis e permite chegar a conclusões de maneira puramente formal, em decorrência exclusiva da sua lógica. Dessa maneira, é possível enquadrar esta pesquisa no método científico dedutivo, uma vez que busca, a

partir da análise e da aplicação dos algoritmos de *Machine Learning*, definir os mais apropriados para a previsão de preços de ativos de renda variável.

Quanto à finalidade, há neste trabalho um exemplo de pesquisa aplicada, uma vez que visa abordar uma problemática real de previsão de preços de ativos de renda variável, além de utilizar dados reais em suas análises.

Em consonância com Chauchick (2012), a abordagem quantitativa trata da captura de evidências para uma pesquisa por meio da mensuração das variáveis, trazendo como uma de suas premissas a objetividade. Além disso, o autor menciona as diferentes técnicas de pesquisas utilizadas nessa abordagem, sendo a *survey*, a modelagem/simulação, o experimento e o quase-experimento. A partir desses conceitos, é possível afirmar que essa pesquisa tem uma natureza predominantemente quantitativa e será pautada na utilização da modelagem/simulação, uma vez que serão analisados indicadores financeiros e *outputs* numéricos referentes aos valores das ações, sendo estes *outputs* provenientes da aplicação de algoritmos de *Machine Learning*.

Na linha de Gil (2002), afirma-se que a pesquisa explicativa tem como cerne identificar os fatores que contribuem ou determinam a ocorrência de fenômenos, a partir disso é possível classificar esta pesquisa como sendo explicativa quanto ao objetivo, uma vez busca investigar as relações de causa e efeito entre as variáveis escolhidas, além de explorar a performances de diferentes combinações de variáveis e algoritmos.

Segundo Lakatos (2003), a documentação indireta é o ponto de partida de todo trabalho científico por tratar do levantamento de dados de variadas fontes que servem de *background* para o campo de interesse, sendo segmentada em fontes primárias e secundárias. É nítido que nesta pesquisa há ambas as fontes, sendo as primárias para os dados financeiros acerca das empresas, e as fontes secundárias para revisão da literatura acerca do tema e suas ramificações.

1.4. Estrutura do trabalho

Este estudo foi estruturado para abordar a utilização de modelos de *Machine Learning* na previsão de preços de ativos de renda variável. É iniciado com uma introdução detalhada, estabelecendo o cenário para o tema em investigação, delineando os objetivos, a relevância do estudo e a metodologia adotada.

No capítulo 2 são estabelecidos os referenciais teóricos que guiarão o desenvolvimento desse trabalho, partindo desde a conceituação acerca da estrutura do mercado financeiro no Brasil até a contextualização acerca das técnicas de *Machine Learning* a serem implementadas.

Posteriormente, no Capítulo 3, é desenvolvida uma extensa revisão da literatura, por meio da exploração extensiva de trabalhos anteriores pertinentes, proporcionando uma sólida base teórica para a construção desse trabalho específico

No Capítulo 4, é buscado explicar de forma minuciosa todas as premissas utilizadas na implementação dos modelos, desde a seleção das empresas a serem avaliadas até o detalhamento das variáveis e algoritmos implementados para realização da análise comparativa.

O Capítulo 5 foca na análise dos resultados obtidos através da implementação dos diferentes modelos em conjunto com as diferentes combinações de variáveis de entrada, aplicados as empresas selecionadas para realizar a análise comparativa do desempenho dos algoritmos e conjuntos de variáveis.

Por fim, temos o Capítulo 6 que aborda os principais conhecimentos desenvolvidos ao longo da construção do trabalho, em conjunto com os *insights* provenientes dos resultados da análise comparativa e sugestões de abordagens futuras acerca do tema dessa dissertação.

2. REFERENCIAL TEÓRICO

Este capítulo apresenta uma base teórica para o desenvolvimento do trabalho, abrangendo conceitos relacionados à composição do mercado financeiro brasileiro, à análise de ativos de renda variável e às técnicas de *Machine Learning*, explorando algoritmos e técnicas avançadas.

2.1. Composição do mercado financeiro brasileiro

O mercado financeiro é um complexo sistema que abrange uma ampla gama de instrumentos financeiros, instituições, regulamentos e participantes, incluindo bancos, fundos de investimento, seguradoras, e investidores individuais e corporativos. Ele permite a alocação eficiente de recursos, facilitando empréstimos, investimentos e a gestão de riscos. A evolução tecnológica, especialmente no campo da informática e telecomunicações, transformou profundamente o mercado financeiro, aumentando a velocidade, eficiência e globalização das transações financeiras, ao mesmo tempo em que introduziu novos desafios e riscos.

Para uma melhor compreensão dos *players* que compõem a estrutura do mercado financeiro brasileiro, temos uma abordagem bastante elucidativa trazida pela B3 (2022), onde temos o Banco Central do Brasil (BACEN) como a autarquia do governo federal que tem por função regular o Sistema Financeiro Nacional, sendo responsável pela definição da política monetária. Seguindo na estrutura regulatória, temos a Comissão de Valores Mobiliários (CVM) que tem por função principal proteger os investidores de práticas fraudulentas. Outro órgão ligado ao governo é a Secretaria do Tesouro Nacional, sendo esta responsável pela disponibilização de títulos públicos ao mercado. Por fim, temos na estrutura governamental a Superintendência de Seguros Privados (SUSEP), órgão responsável pela fiscalização do mercado de seguros, resseguros e previdência no país.

Ainda na composição trazida pela B3 (2022), há na estrutura privada a própria B3 que é a única bolsa de valores em operação atualmente no Brasil, sendo responsável por gerir o mercado de compra e venda de valores mobiliários. Essa estrutura é composta também pelos bancos, responsáveis pelo oferecimento de serviços financeiros em geral, pelas corretoras, as quais possuem o papel de auxiliar na realização de investimento por pessoas físicas, podendo essas terem relação com os bancos ou não, existindo ainda as Gestoras de Recursos, responsáveis pela gestão ativa do patrimônio de seus clientes através de fundos de investimento e as Cooperativas de Crédito, que atuam mais ativamente na concessão de crédito para pessoas físicas. Por fim, existem as empresas

de capital aberto, as quais serão o principal objeto de estudo desse trabalho, e que são caracterizadas pelo caráter público de suas ações, ou seja, qualquer pessoa ou empresa que esteja apta a negociar no mercado financeiro pode adquirir uma participação dessas empresas.

Todas as instituições citadas anteriormente fazem parte da composição do Sistema Financeiro Nacional (SFN), o qual pode ser definido como um conjunto de mecanismos, instrumentos e instituições responsáveis por assegurar o direcionamento de recursos para investimento de acordo com B3 (2022). Sendo este composto por seis principais mercados, conforme ilustrado na Figura 1.

Figura 1 – Composição do Sistema Financeiro Nacional



Fonte: B3 (2024)

De acordo com o “Guia Por Dentro da B3” (B3, 2024), o mercado monetário foca no controle da oferta de moeda para garantir a liquidez, sendo regulado pelo Banco Central do Brasil (BACEN) através da política monetária. Seguindo nesse viés, o mercado de crédito é caracterizado pela intermediação de recursos financeiros para agentes deficitários, sendo este também parte do escopo de atuação do BACEN. O mercado de câmbio, por sua vez, trata da negociação de moedas estrangeiras, também sendo alvo da regulação do BACEN. Composto ainda o SFN, temos o mercado de seguros e previdência, focado em instrumentos de transferência de mitigação de riscos, ficando sua fiscalização sobre responsabilidade majoritária da SUSEP.

Por fim, existem dois mercados os quais estão mais relacionados ao objeto de estudo desse trabalho, o mercado de capitais, responsável por direcionar recursos para a capitalização de

empresas de capital aberto, sendo neste onde ocorre a definição do valor de mercado de um determinado ativo de acordo com as operações de compra e venda por parte dos investidores, e o mercado de derivativos, o qual tem por função a negociação de instrumentos para mitigação de riscos, sendo estes contratos atrelados a um bem subjacente que pode ser proveniente do mercado de capitais, como a ação de uma empresa.

2.2. Análise de ativos de renda variável

Um conceito interessante trazido por M. C. Thomsett (2015) *apud* Beyaz *et al.* (2018) é que o preço de uma ação é determinado a partir da relação entre oferta e demanda por aquele ativo, sendo estas guiadas pelo interesse em comprar ou vender por parte dos *players* do mercado.

A análise de ativos de renda variável é um campo fundamental para compreender e prever os movimentos de preços no mercado financeiro. Os estudos de Fabozzi *et al.* (2007) e Malta, Lessa e De Camargos (2016) destacam a importância da análise fundamentalista na avaliação de ativos. A análise fundamentalista é composta por fatores econômicos, financeiros e setoriais das empresas, como indicadores financeiros, balanços patrimoniais e perspectivas de mercado.

Em paralelo, há o viés técnico que, conforme Beyaz *et al.* (2018), busca identificar sinais de compra e venda com base no uso de técnicas estatísticas para análise de padrões de comportamento do mercado, como suportes, resistências, médias móveis e indicadores de *momentum*. A análise técnica é amplamente utilizada por *traders* e investidores para tomar decisões de curto prazo e identificar oportunidades de lucro no mercado financeiro.

Damodaran (2012) apresenta diferentes modelos para a elaboração do *valuation* de uma empresa, ou seja, para determinar o seu valor intrínseco a partir dos fundamentos do negócio. Um desses modelos consiste na determinação da projeção dos fluxos de caixa futuros de uma empresa a valor presente por meio do desconto a uma taxa que reflita o risco da companhia, sendo este um dos métodos mais populares no segmento fundamentalista, o “fluxo de caixa descontando”. Em outro modelo, o *valuation* por múltiplos, empresas de um mesmo setor são comparadas por meio de múltiplos como Preço / Lucro, *Return on Equity (ROE)*, entre outros, sendo um método bastante relevante para realização de *stock picking*, dentro de um setor específico. Sendo *stock picking* o processo de escolha de ações individuais para composição do portfólio de investimentos com o objetivo de obter um retorno superior ao retorno do mercado,

Ao partir para uma imersão no viés técnico, é contextualizado em Lemos (2018) que um dos pilares da análise técnica são os gráficos de preços, que incluem formatos como linhas, barras e *candles*, que mostram o comportamento dos preços de um ativo ao longo do tempo. Além disso, são

empregados indicadores técnicos, calculados a partir de dados de preços e/ou volumes. Esses indicadores ajudam a identificar tendências, condições de compra ou venda, e outros padrões. Alguns exemplos incluem Médias Móveis, *Relative Strength Index (RSI)*, *Moving Average Convergence Divergence (MACD)* e Bandas de Bollinger.

Além disso, Lemos (2018) enfatiza que a análise técnica também envolve a identificação de padrões gráficos, como padrões de cabeça e ombros, triângulos, bandeiras e outros. Acredita-se que esses padrões possam indicar potenciais reversões de tendência ou a continuação de movimentos previamente observados. O volume de negociação é frequentemente considerado um indicador crucial na análise técnica, onde mudanças significativas no podem ser indicativos de pontos de virada no mercado, proporcionando *insights* valiosos aos analistas.

Por fim, existem ainda os conceitos de suporte e resistência, onde níveis de preços representam áreas onde se acredita que um ativo pode encontrar dificuldades para subir (resistência) ou cair (suporte). A análise técnica busca identificar esses níveis como referências importantes para tomadas de decisão de compra ou venda.

2.3. Técnicas de Machine Learning

De acordo com Domingos (2012), *Machine Learning* é uma subárea da inteligência artificial que se concentra na criação de algoritmos e modelos com a capacidade de discernir padrões a partir de dados. Essa habilidade permite que sistemas realizem previsões, tomem decisões e identifiquem padrões sem a necessidade de programação explícita para essas tarefas. As aplicações relacionadas a *Machine Learning* são vastas e abrangem diversos campos, como reconhecimento de padrões, processamento de linguagem natural, visão computacional, previsão de valores estocásticos, entre outras.

Além disso, há nas abordagens destacadas por Domingos (2012), o aprendizado supervisionado como uma das estratégias mais utilizadas para algoritmos de *Machine Learning*. Nesse contexto, os algoritmos são treinados por meio de conjuntos de dados que possuem rótulos, ou seja, as respostas corretas associadas a eles. O objetivo é capacitar o algoritmo a aprender e generalizar a partir desses exemplos rotulados, permitindo que ele faça previsões ou classificações precisas em novos dados.

Em contrapartida, o aprendizado não supervisionado também é uma abordagem crucial, enfocando a descoberta de padrões e estruturas em conjuntos de dados não rotulados. Isso é especialmente útil para tarefas como agrupamento (*clustering*) e redução de dimensionalidade, onde a identificação de relações subjacentes nos dados é o objetivo principal.

A seguir são apresentados dois *Machine Learning*, fundamentais para o desenvolvimento deste trabalho.

2.3.1. Support Vector Regression (SVR)

O *Support Vector Regression (SVR)* é destacado por Wang (2005) como uma metodologia de aprendizado de máquina de grande eficácia, empregada amplamente em tarefas de classificação, regressão e outros tipos de aprendizado supervisionado. A aplicabilidade do *SVR* na previsão de preços de ações é ressaltada pela sua habilidade em modelar relações complexas e não lineares entre variáveis de mercado e preços futuros, visando a acurácia nas previsões.

A essência do *SVR* reside na identificação do hiperplano ideal que distingue as categorias nos dados de treino, como os movimentos ascendentes ou descendentes dos preços das ações, otimizando a margem entre os vetores de suporte, ou seja, os dados mais próximos ao hiperplano. Tal estratégia assegura a robustez do modelo e sua capacidade de generalização para dados não observados anteriormente.

No contexto financeiro, caracterizado pela não linearidade e alta dimensionalidade dos dados, o *SVR* se mostra particularmente vantajoso, adaptando-se a essas complexidades através do uso de funções kernel, como as polinomiais, *Radial Basis Function (RBF)* ou sigmoidais. Essas funções permitem a operação do *SVR* em espaços de alta dimensão, facilitando a separação linear dos dados sem a necessidade de computar explicitamente as transformações espaciais.

Para a previsão de preços no mercado de ações utilizando o *SVR*, é comum a adoção de indicadores técnicos e fundamentalistas como *inputs*, adaptando o problema para uma abordagem de regressão ou classificação. A eficácia do *SVR* nesse contexto é fortemente influenciada pela seleção adequada de características, ajuste de parâmetros e a implementação de validação cruzada para prevenir o *overfitting* e assegurar a generalização do modelo.

Conforme explorado ainda por Wang (2005), a aplicação do *SVR* na previsão de preços de ações tem sido validada por diversos estudos, que evidenciam sua capacidade de entender a dinâmica do mercado e contribuir para estratégias de investimento eficazes. Entretanto, é crucial reconhecer as limitações inerentes ao método, como a sensibilidade aos parâmetros do modelo e a necessidade de dados de alta qualidade, para maximizar seu potencial preditivo no âmbito financeiro.

2.3.2. Random Forest

Em Polamuri *et al.* (2019), é apresentada a técnica do *Random Forest* aplicada à previsão de preços de ações, destacando sua eficácia em capturar a dinâmica complexa dos mercados financeiros. O estudo ressalta como o *Random Forest*, ao construir múltiplas árvores de decisão com amostras e características aleatórias, oferece robustez e diminui o risco de *overfitting*, sendo uma abordagem interessante para a análise indicadores financeiros e previsão dos movimentos de preços no mercado de ações.

Na previsão de preços de ações, o *Random Forest* analisa diversos indicadores técnicos e fundamentalistas, como volume de negociação, médias móveis, índices de força relativa, entre outros, para prever a direção dos movimentos dos preços ou os próprios valores futuros. Cada árvore da floresta fornece uma predição, e a previsão final do modelo é determinada pela média (no caso de regressão) ou pela maioria dos votos (no caso de classificação) das predições de todas as árvores.

A principal vantagem do *Random Forest* na previsão de preços de ações é sua capacidade de lidar com grandes conjuntos de dados com múltiplas variáveis de entrada, capturando interações complexas entre elas sem a necessidade de uma especificação de modelo detalhada. Além disso, o método é menos suscetível a *overfitting* em comparação com uma única árvore de decisão, devido ao processo de agregação das previsões das árvores individuais.

Apesar de suas vantagens, o uso do *Random Forest* na previsão de preços de ações também apresenta desafios, como a interpretação dos resultados, dado que a lógica por trás das predições de uma floresta de árvores pode ser menos transparente do que modelos mais simples.

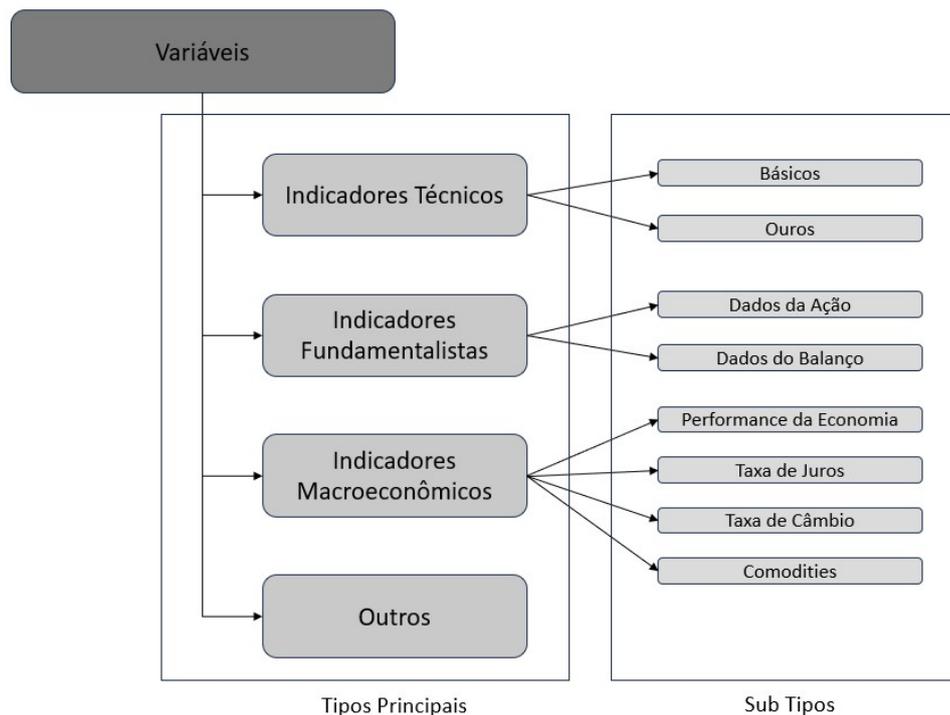
Em resumo, o *Random Forest* oferece uma abordagem poderosa e flexível para a previsão de preços de ações, com potencial para capturar a complexidade dos mercados financeiros e fornecer *insights* valiosos para investidores e analistas. No entanto, a eficácia do método depende de uma implementação cuidadosa e da consideração de suas limitações.

3. REVISÃO DA LITERATURA

A aplicação de técnicas de *Machine Learning* na previsão de preços de ativos de renda variável tem se mostrado promissora. Vários estudos exploraram diferentes algoritmos e abordagens nessa e em outras áreas que serviram de base para o desenvolvimento deste trabalho.

Há em Kumbure *et al.* (2022) uma extensa revisão da literatura acerca da previsão de preços de ativos de renda variável utilizando técnicas de *Machine Learning*, o artigo em questão traz uma visão dos diferentes indicadores e modelos utilizados em 138 artigos publicados desde o ano 2000. O *framework* representado pela Figura 2 auxilia no entendimento das diferentes variáveis utilizadas para previsão de preços, abordando a distribuição das 2.173 variáveis tratadas nos artigos analisados em quatro principais tipos, dos quais dois serão utilizados nesse trabalho, tanto de forma isolada quanto combinada para efeito comparativo.

Figura 2 - Categorias das variáveis utilizadas para previsão de preços de ações (adaptado de Kumbure et al, 2022)



Fonte: Elaborado pelo autor, baseado em Kumbure *et al.*, 2022

Em Mehas *et al.* (2020), há a aplicação e comparação dos resultados de diferentes técnicas como *Random Forest*, *Gradient Boosting* e *Support Vector Regression*, na previsão do preço de fechamento de ações, concluindo que cada um destes algoritmos possui capacidades específicas e a escolha do modelo mais adequado depende do contexto e dos dados disponíveis.

Acerca da técnica do *Support Vector Regression*, é observado que esta é muito eficiente em lidar com problemas de regressão complexos e não lineares (Awad e Khanna 2015). O *Support Vector Regression* utiliza os mesmos princípios do *Support Vector Machine* que é amplamente utilizado em problemas de classificação. O estudo de Jain, Rahman & Kulkarni (2007) explorou o uso do *Support Vector Regression* no desenvolvimento de um "soft sensor" para uma coluna de destilação em batelada, a qual busca encontrar um hiperplano no espaço de alta dimensão que maximize a margem entre os dados de treinamento e o hiperplano, enquanto mantém as previsões dentro de uma faixa de tolerância definida. Essa abordagem permite capturar relações não lineares entre as variáveis de entrada e saída.

No tocante ao *Random Forest*, Harikrishnan *et al* (2021) o utiliza para prever preços de ações, treinado com dados históricos de preços e indicadores técnicos. Ao combinar as previsões das árvores de decisão, o *Random Forest* captura padrões e tendências nos dados de forma robusta, tendo como vantagens a capacidade de lidar com problemas de regressão e classificação, ser menos suscetível a *overfitting* e possibilitar a avaliação da importância das variáveis utilizadas na previsão.

Em Fabozzi, Focardi e Jonas (2007), foi avaliada a adoção de técnicas e abordagens quantitativas, como modelos de precificação de ativos, otimização de portfólio, análise de fatores e estratégias de negociação sistemática por meio de pesquisas com gestores de fundos de investimento e profissionais do setor financeiro, permitindo incorporar aprendizados acerca do funcionamento das finanças quantitativas no mercado.

Já em Beyaz *et al.* (2018), é possível desenvolver uma base de aprendizado envolvendo a comparação do uso de indicadores técnicos e fundamentalistas na previsão do preço de ativos de renda variável, sendo um dos temas chaves desta pesquisa, enquanto em Malta, Lessa e de Camargos (2016) temos uma abordagem mais específica dos indicadores fundamentalistas no contexto de empresas brasileiras. De forma complementar, em Oliveira, Rêgo & Diniz (2019), temos a aplicação de técnicas de *Machine Learning* e *Data Science* para previsão de valores de aquisições governamentais, a qual possui similaridades com esta pesquisa.

Enquanto em Santos (2020) ocorre a aplicação dos métodos no contexto que se dará esta pesquisa, porém restrita a ativos da bolsa de valores brasileira (B3), em Mehar (2020) existe a aplicação mais específica dos algoritmos de redes neurais e *Support Vector Machine*. Além disso, é possível observar em Brunhuemer (2022) uma aplicação voltada a classificação de estratégias de *short straddle*, por fim, é encontrada uma abordagem mais geral do tema na análise de Harikrishnan (2021), que busca avaliar diferentes modelos aplicados na previsão dos preços de ações.

No entanto, é em Awad e Khanna (2015) que pode-se extrair a base teórica para conceituação do *SVR*, enquanto em Jain, Rahman & Kulkarni (2007) existe a aplicação deste algoritmo para o

desenvolvimento de um *soft sensor* (sensor virtual) para uma coluna de destilação em batelada com o objetivo de estimar as variáveis do processo de difícil mensuração com base nas variáveis do processo de mensuração mais simples.

A aplicação de técnicas de Machine Learning na previsão de preços de ativos de renda variável representa um campo de estudo vasto e dinâmico, que tem se expandido e evoluído consideravelmente ao longo dos anos. A revisão de literatura realizada por Kumbure *et al.* (2022) e os diversos estudos citados ilustram a diversidade de abordagens e algoritmos explorados, desde *Random Forest* e *Gradient Boosting* até *Support Vector Regression*, cada um com suas particularidades e aplicações específicas.

Sendo destacado ainda a eficácia destas técnicas em capturar relações complexas e não lineares entre variáveis, juntamente com a capacidade de processar grandes volumes de dados, ilustrando o potencial dos modelos de *Machine Learning* em fornecer *insights* valiosos e previsões precisas no contexto dos mercados financeiros.

Por tanto, o *SVR* foi selecionado para implementação por, de acordo com Awad e Khanna (2015), apresentar elevada eficiência em lidar com dados de alta dimensionalidade e complexidade, além de ser bastante adequado para modelar relações não-lineares bastante presentes em dados financeiros.

De forma similar, a seleção do *Random Forest* ocorreu devido a sua robustez e capacidade de redução do *overfitting* conforme ilustrado por Polamuri *et al.* (2019), além da eficácia deste modelo na previsão de preços de ativos de renda variável apresentada em Vijn *et al.* (2020) e Santos (2020).

4. DESENVOLVIMENTO DOS MODELOS

Neste capítulo, haverá o detalhamento acerca de conceitos importantes aplicados no desenvolvimento do modelo para obtenção dos resultados a serem discutidos nos próximos capítulos.

Inicialmente, apresenta-se o detalhamento acerca das empresas que farão parte da análise comparativa das diferentes combinações entre modelos e indicadores em conjunto com as motivações por trás da escolha de a cada uma delas.

Posteriormente, para um melhor entendimento quanto à implementação e à análise desenvolvida, serão abordados cada um dos indicadores técnicos e fundamentalistas de forma individualizada em conjunto com as ferramentas utilizadas para extração e tratamento dos dados.

Por fim, será desenvolvida uma breve conceituação acerca do funcionamento de cada um dos algoritmos de *Machine Learning* a serem implementados, buscando elucidar como cada um é aplicado e as ferramentas utilizadas para a sua aplicação, em conjunto com o detalhamento das comparações a serem realizadas através das diferentes combinações entre tipos de indicadores e modelos.

4.1. Empresas analisadas

Para o desenvolvimento da análise comparativa entre os modelos e as diferentes combinações de tipos de variáveis, foram selecionadas duas empresas de setores e nacionalidades distintas. Essa seleção é estrategicamente valiosa devido à diversidade das condições de mercado e dos ambientes regulatórios em que essas empresas operam.

Essa abordagem permite testar a robustez e a adaptabilidade dos modelos em cenários variados, o que é crucial para verificar sua eficácia em prever tendências em contextos distintos. A diversidade setorial expõe os modelos a dinâmicas de mercado únicas, incluindo diferentes padrões de risco, oportunidades e influências de fatores econômicos, políticos e sociais

A inclusão de empresas de várias nacionalidades adiciona outra camada de complexidade, introduzindo variáveis como diferenças regulatórias, condições econômicas e tendências de consumo específicas de cada país. Essa heterogeneidade ajuda a avaliar como os modelos podem ser ajustados ou generalizados para refletir as peculiaridades de diferentes ambientes econômicos e regulatórios.

Essa variedade de setores e nacionalidades, impacta também as variáveis utilizadas que podem ter diferentes graus de relevância e interpretação em cada setor e localidade. Por exemplo, indicadores como o Preço / Lucro (P/L) e o Retorno sobre o Patrimônio Líquido (*ROE*) podem ser influenciados pela estrutura de capital e pelas expectativas de crescimento específicas de cada setor, enquanto indicadores técnicos como o *RSI* e o *MACD* podem refletir diferentes padrões de volatilidade e comportamento de preços em mercados distintos.

Portanto, a análise comparativa utilizando empresas de setores e nacionalidades diferentes não apenas testa a flexibilidade e adaptabilidade dos modelos de *Machine Learning* diante de variadas condições de mercado, mas também proporciona *insights* valiosos sobre como diferentes indicadores podem ser utilizados para capturar nuances específicas relacionadas ao setor e à localidade das empresas analisadas. Essa abordagem holística e diversificada contribui significativamente para o desenvolvimento de estratégias de investimento mais informadas e adaptadas à complexidade do mercado global.

A primeira empresa selecionada para compor a análise foi a Petrobras, Petróleo Brasileiro S.A., destacando-se como uma gigante no setor energético, com atividades que se estendem da exploração e produção de petróleo (*upstream*) ao refino, comercialização e transporte de petróleo e derivados (*downstream*), tanto no Brasil quanto no cenário internacional. A empresa possui como controlador e acionista majoritário o governo brasileiro, ocupa uma posição central na indústria energética brasileira e exerce uma influência notável no mercado global de petróleo.

Figura 3 – Logomarca Petrobras



Fonte: <https://www.petrobras.com.br/>

Ademais, a expertise da empresa em exploração e produção em águas profundas e ultraprofundas é reconhecida mundialmente, especialmente devido às suas inovações tecnológicas desenvolvidas para a exploração do Pré-sal, uma das maiores descobertas petrolíferas das últimas décadas. Embora seu foco principal seja o petróleo, a Petrobras também tem iniciativas voltadas para as energias renováveis, buscando diversificar sua matriz energética.

Ao fazer uma análise do setor de Óleo e Gás, no qual a Petrobras está inserida, é possível observar que ela compete tanto no mercado nacional, dominando o cenário brasileiro, quanto no

internacional, enfrentando grandes empresas petrolíferas estatais e multinacionais como Saudi Aramco, ExxonMobil e BP (British Petroleum). Além disso, tem como principais riscos associados a volatilidade dos preços do petróleo, desafios regulatórios, impactos ambientais de derramamentos de óleo e riscos políticos devido ao seu controle estatal, todos os quais podem influenciar suas estratégias e operações. Destaca-se ainda que o setor de Óleo e Gás requer intensivos investimentos em capital e tecnologia, especialmente para projetos em locais remotos e em águas profundas, o que pode representar uma grande barreira de entrada a novos entrantes.

A escolha da Petrobras para a comparação dos modelos de *Machine Learning* é justificada por sua representatividade significativa no mercado de Óleo e Gás, além da rica disponibilidade de dados e a diversidade de desafios que enfrenta. A análise da companhia pode oferecer *insights* valiosos sobre como a performance dos modelos aplicados em empresas de grande porte e setores voláteis, fornecendo uma oportunidade de entender como diferentes indicadores técnicos e fundamentalistas influenciam as previsões de preços de ações em contextos complexos e dinâmicos.

A outra empresa selecionada para compor a análise é a Amazon, uma das empresas mais influentes e inovadoras do mundo, originalmente fundada como uma livraria online por Jeff Bezos em 1994. Desde então, evoluiu para se tornar uma gigante do *e-commerce*, oferecendo uma vasta gama de produtos e serviços, que vão desde produtos de consumo até tecnologia de computação em nuvem através da sua subsidiária *Amazon Web Services* (AWS). A Amazon também está envolvida em áreas como produção de conteúdo, eletrônicos de consumo, e no setor de varejo alimentício através da aquisição da Whole Foods.

Figura 4 – Logomarca Amazon



Fonte: <https://www.aboutamazon.com/>

A empresa opera em um setor de varejo online altamente competitivo e em rápida evolução, caracterizado pela conveniência, velocidade de entrega e personalização da experiência do cliente. Além disso, a AWS coloca a Amazon como uma líder no mercado de serviços de computação em nuvem, competindo com gigantes como Microsoft Azure e Google Cloud Platform.

A companhia possui como principais competidores no segmento de *e-commerce* o eBay, Walmart, Alibaba e uma série de varejistas especializados em várias regiões. No segmento de computação em nuvem, além da Microsoft e do Google, enfrenta concorrentes como IBM e Oracle.

No setor de conteúdo digital e dispositivos, a Amazon compete com Apple, Google e uma variedade de serviços de streaming de mídia, como Netflix e Disney+.

Ao analisar os riscos associados à operação da Amazon, é fundamental destacar a intensa concorrência em todos os segmentos de mercado em que atua, exigindo um foco em inovação por meio de grandes investimentos. A empresa também enfrenta desafios regulatórios em diversas jurisdições, dada a sua vasta operação e preocupações antitruste. Outros riscos incluem questões de privacidade e segurança de dados, especialmente relacionadas à AWS e à crescente dependência da sociedade em serviços de computação em nuvem.

Por fim, a utilização da Amazon na análise comparativa é justificada pela sua complexidade operacional e pela diversidade de dados gerados por suas múltiplas linhas de negócios. A análise da companhia irá fornecer *insights* acerca do desempenho dos modelos em capturar e prever as dinâmicas de mercado em empresas que operam em múltiplos setores, especialmente em ambientes altamente competitivos e tecnologicamente avançados.

4.2. Tipos de indicadores

Nos modelos desenvolvidos, foram utilizados dois tipos de indicadores como variáveis para os algoritmos de aprendizagem de máquina, os indicadores fundamentalistas e os indicadores técnicos. O primeiro deles é composto por métricas que permitem a avaliação da saúde financeira da empresa e do seu potencial de crescimento por meio da análise de fatores como lucratividade, endividamento e fluxo de caixa.

Enquanto os indicadores técnicos funcionam como ferramentas matemáticas calculadas a partir de estatística aplicada aos dados de preço e/ou volume de negociação de uma ação, sendo bastante utilizados para a identificação de tendências em relação ao preço da ação.

4.3. Indicadores fundamentalistas

No espectro fundamentalista o primeiro indicador selecionado foi o Preço/Lucro (P/L), também conhecido como múltiplo de lucros, é uma métrica amplamente utilizada para avaliar o valor relativo de uma ação e para a realização de comparações entre empresas de um mesmo setor. O P/L é calculado pela equação:

$$\frac{P}{L} = \frac{\text{Preço por Ação}}{\text{Lucro por Ação}} \quad (1)$$

É importante destacar ainda que a aplicação do indicador P/L varia significativamente entre diferentes setores, devido às distintas características de crescimento, margens de lucro e estágios de desenvolvimento das empresas em cada setor. O P/L é especialmente influente na comparação de empresas dentro do mesmo setor, pois permite uma avaliação relativa da valorização de empresas que compartilham um ambiente de mercado semelhante.

Nos setores tradicionais, como bens de consumo, finanças e indústrias, onde as empresas geralmente apresentam crescimento estável e modelos de negócios bem estabelecidos, o P/L é frequentemente utilizado para avaliar se as ações estão subvalorizadas ou sobrevalorizadas em relação aos seus pares. Um P/L significativamente mais alto ou mais baixo que a média do setor pode indicar uma sobrevalorização ou subvalorização, respectivamente, assumindo que as empresas tenham estruturas de capital semelhantes e estejam em estágios comparáveis de crescimento.

No setor de tecnologia, no entanto, a utilização do P/L pode ser menos intuitiva ou até mesmo enganosa pelas seguintes razões:

- **Alto Crescimento:** Muitas empresas de tecnologia estão em estágios de crescimento rápido, reinvestindo a maior parte de seus lucros para expandir o negócio. Isso pode resultar em lucros relativamente baixos ou até mesmo em perdas, distorcendo o cálculo do P/L;
- **Expectativas Futuras:** O valor das empresas de tecnologia muitas vezes reflete expectativas de crescimento futuro, não necessariamente os lucros atuais. Isso pode levar a P/Ls extremamente altos, que não fazem sentido quando comparados a setores mais estáveis;
- **Inovação e disrupção:** Empresas de tecnologia muitas vezes operam em mercados altamente voláteis e competitivos, com inovações constantes. O sucesso financeiro de longo prazo pode depender da capacidade da empresa de continuar inovando, o que é difícil de quantificar no P/L;
- **Modelos de Negócios Diversificados:** Muitas empresas de tecnologia têm múltiplas fontes de receita e modelos de negócios complexos que podem não ser diretamente comparáveis, tornando as comparações de P/L dentro do setor menos significativas.

Por essas razões, enquanto o P/L pode ser uma ferramenta útil para avaliar empresas em setores mais tradicionais, analistas e investidores muitas vezes buscam outros indicadores, como

crescimento de receita, margens de lucro, fluxo de caixa livre, entre outros, para avaliar empresas de tecnologia. Isso não significa que o P/L seja totalmente irrelevante no setor de tecnologia, mas ele precisa ser considerado em conjunto com uma gama mais ampla de fatores para gerar *insights* mais assertivos acerca do negócio.

Dessa maneira, o P/L é um indicador bastante relevante para integrar as variáveis dos modelos de *Machine Learning*, uma vez que fornece insights concisos sobre a relação preço-lucro, facilitando a identificação de padrões de preços em relação aos fundamentos financeiros das empresas. Sendo crucial para previsão de movimentos futuros de preços e tornando-o uma componente essencial na modelagem preditiva de desempenho financeiro.

Um outro indicador de viés fundamentalista utilizado como variável foi o Retorno Sobre o Patrimônio Líquido (*ROE*, do inglês *Return on Equity*), métrica financeira que mede a capacidade de uma empresa gerar lucro a partir do seu patrimônio líquido, o qual consiste na diferença entre o ativo e o passivo de uma determinada empresa, sendo o *ROE* calculado pela equação:

$$ROE = \left(\frac{\text{Lucro Líquido}}{\text{Patrimônio Líquido}} \right) \times 100 \quad (2)$$

De forma direta, o *ROE* é um indicador poderoso que reflete a eficiência da gestão da empresa em gerar valor para os acionistas, fornecendo insights valiosos para a análise do desempenho futuro de uma ação.

Um *ROE* elevado não apenas destaca a capacidade da empresa de reinvestir seus lucros de forma eficaz, resultando em crescimento acelerado e sustentável a longo prazo, mas também sinaliza um equilíbrio favorável entre risco e retorno. Esse equilíbrio é essencial para modelos preditivos que visam otimizar as estratégias de investimento, tornando este indicador um componente crítico na avaliação da atratividade e do potencial de uma empresa.

Por fim, é importante destacar a utilização do indicador Dívida / LAJIDA (Lucro antes de juros, impostos, depreciação e amortização) uma razão financeira que compara a dívida total de uma empresa ao seu LAJIDA, sendo calculado pela equação:

$$\text{Dívida/LAJIDA} = \frac{\text{Dívida Total}}{\text{LAJIDA}} \quad (3)$$

A Dívida Total inclui todas as obrigações da empresa, como empréstimos bancários, títulos de dívida, notas promissórias e outras formas de dívida, enquanto o LAJIDA oferece uma visão

mais clara dos ganhos operacionais da empresa e é uma métrica bastante utilizada dentro do mercado financeiro.

A combinação dessas duas métricas fornece uma medida da alavancagem financeiro do negócio, quanto mais alta essa alavancagem maior o risco financeiro. Além disso, permite a avaliação da capacidade de pagamento da dívida, sendo muito utilizada na avaliação dos credores para definição de *ratings* de crédito e taxa de juros nas emissões de dívidas, empresas com maior relação Dívida / LAJIDA irão ter um custo maior na tomada de crédito no mercado.

Dessa forma, o indicador Dívida / LAJIDA é essencial na avaliação do risco financeiro, pois oferece insights sobre a alavancagem e sustentabilidade financeira de uma empresa, destacando a relação entre sua dívida e eficiência operacional medidos pelo LAJIDA. Este indicador é particularmente valioso em modelos de *Machine Learning*, pois ao integrá-lo, é possível enriquecer a análise considerando tanto a estrutura de capital da empresa quanto sua capacidade de administrar obrigações financeiras, elementos cruciais para antecipar a performance futura de suas ações no mercado.

4.4. Indicadores técnicos

O primeiro indicador a ser selecionado foi o *Relative Strength Index (RSI)* que, de maneira simplificada, busca medir a velocidade e a mudança de movimentos de preços, sendo um oscilador de *momentum* que varia de 0 a 100. Este indicador é calculado de acordo com a equação:

$$RSI = 100 - \frac{100}{1 + \frac{MG}{MP}} \quad (4)$$

Onde:

MG – média aritmética dos preços de fechamento em alta

MP – média aritmética dos preços de fechamento em baixa

O *RSI* busca identificar sinais de que um ativo esteja sobrecomprado, ou seja, é esperada uma correção dos preços no futuro, ou de que esteja sobrevendido, podendo indicar uma tendência de valorização em breve. Este indicador foi selecionado para compor os modelos devido à sua popularidade e eficácia em mensurar o *momentum* do mercado e identificar possíveis pontos de reversão. Este indicador fornece informações úteis sobre as tendências do mercado e as pressões de

compra e venda, sendo bastante valioso para modelos preditivos que tem por objetivo prever movimentos futuros de preços.

Outro indicador técnico utilizado como variável nos modelos foi a Média Móvel Simples (MMS) de 21 períodos, sendo calculado como a média aritmética dos preços de um ativo durante os últimos 21 períodos. Sendo calculado por meio da seguinte equação:

$$MMS_n = \frac{\sum_{i=1}^n x_i}{n} \quad (5)$$

Onde:

n – número de períodos

x – preço de fechamento

O MMS é comumente utilizado na análise técnica para avaliar a direção da tendência de um ativo, indicando uma tendência de alta quando o preço está acima da média e de baixa quando está abaixo. No entanto, devido à sua natureza de ponderação igual, o MMS pode reagir lentamente às mudanças recentes de preços.

Para contornar esse efeito, foi utilizada também a Média Móvel Exponencial (MME) de 7 períodos, sendo este um indicador que calcula a média de preços de um ativo priorizando os dados mais recentes através de um fator de suavização, tornando-o mais responsivo às últimas mudanças de preço em comparação ao MMS. Este indicador é bastante utilizado na análise técnica para detectar rapidamente as tendências ou reversões do mercado e é descrito pela equação a seguir:

$$MME_n = x_i \times \left(\frac{2}{n + 1} \right) + MME_{n-1} \times \left(1 - \frac{2}{n + 1} \right) \quad (6)$$

Onde:

n – número de períodos

x – preço de fechamento

Essa combinação de médias móveis permite capturar aspectos distintos do movimento de preço. Enquanto o MMS auxilia o modelo no entendimento da tendência mais a longo prazo, o MME fornece informação acerca do preço do ativo no curto prazo. Tal combinação permite que o modelo seja capaz de identificar padrões e tendências no comportamento dos preços que podem não ser evidentes ao olhar apenas para os preços de fechamento diários isolados.

Por último, foi selecionado o *On-Balance Volume (OBV)*, um indicador técnico que, diferentemente dos anteriores, utiliza o volume de negociação para prever mudanças no preço do ativo. O *OBV* é iniciado com um número arbitrário (geralmente o volume do primeiro dia do período analisado) e então ajustado diariamente, dependendo da movimentação do preço do ativo de acordo com a seguinte equação:

$$OBV = OBV_{\text{período anterior}} + \begin{cases} V & \text{se } F_N > F_{N-1} \\ 0 & \text{se } F_N = F_{N-1} \\ -V & \text{se } F_N < F_{N-1} \end{cases} \quad (7)$$

Onde:

V – volume de negociação

F – preço de fechamento

N – período

Como variável para modelos de *Machine Learning*, o *OBV* pode ser um indicador bastante valioso, pois é capaz de capturar informações não refletidas apenas pelo preço. Quando o volume e o preço se movem na mesma direção, isso pode reforçar a presença de uma tendência forte.

Por fim, na Tabela 1 é feita uma síntese de todos os indicadores e suas respectivas fórmulas para auxiliar no entendimento geral dos indicadores a serem utilizados.

Tabela 1 – Indicadores selecionados para utilização nos modelos de Machine Learning

Indicadores Fundamentalistas		Indicadores Técnicos	
Notação	Fórmula	Notação	Fórmula
P/L	$P/L = \frac{\text{Preço por Ação}}{\text{Lucro por Ação}}$	RSI	$RSI = 100 - \frac{100}{1 + \frac{MG}{MP}}$
ROE	$ROE = \left(\frac{\text{Lucro Líquido}}{\text{Patrimônio Líquido}} \right) * 100$	MMS	$MMS_n = \frac{\sum_{i=1}^n x_i}{n}$
Dívida/LAJIDA	$Dívida / LAJIDA = \frac{Dívida Total}{LAJIDA}$	MME7	$MME_n = x_i * \left(\frac{2}{n+1} \right) + MME_{n-1} * \left(1 - \frac{2}{n-1} \right)$
		OBV	$OBV = OBV_{\text{período anterior}} + \begin{cases} V & \text{se } F_N > F_{N-1} \\ 0 & \text{se } F_N = F_{N-1} \\ -V & \text{se } F_N < F_{N-1} \end{cases}$

Fonte: Elaborado pelo autor (2024)

4.5. Support Vector Regression (SVR)

Um dos algoritmos de *Machine Learning* utilizado nesse estudo para a realização da análise comparativa foi o *SVR*, tido como uma extensão do *Support Vector Machine (SVM)*, porém com foco na previsão de valores contínuos. O objetivo principal do *SVR* é encontrar o hiperplano que melhor se ajusta ao conjunto de dados incorporando parâmetros em relação ao *SVM* para tratar da natureza contínua dos dados.

De forma resumida o algoritmo *SVR*, busca encontrar um hiperplano que melhor se ajuste aos dados dentro de uma margem de tolerância definida pelo parâmetro “*épsilon*” (ϵ), onde os erros dentro dessa margem não são penalizados, utilizando vetores de suporte, pontos de dados próximos ao hiperplano, para definir sua posição e orientação, enquanto emprega a função de perda “*ε-insensitive*” que ignora erros menores que ϵ , tornando o modelo robusto a pequenos erros.

É importante destacar que o *SVR* pode utilizar a técnica matemática do *kernel trick*, para modelar relações não lineares, sendo a função kernel *Radial Basis Function (RBF)* uma das mais comuns utilizadas para essa finalidade.

Por fim, existem ainda dois parâmetros importantes no algoritmo *SVR*, o parâmetro “*C*” e o parâmetro “*gamma*” que juntos ao “*épsilon*” (ϵ) desempenham um papel fundamental na determinação dos ajustes do modelo e da sua capacidade de generalização para diferentes conjuntos de dados.

Sendo “*C*” um hiperparâmetro de regularização que determina o nível de regularização a ser aplicada, ou seja, controla o *trade-off* entre o nível de complexidade do modelo e a capacidade de generalização do modelo. E “*gamma*” responsável por definir o nível de influência que um único ponto da amostra de treinamento pode alcançar.

4.6. Random Forest

O outro algoritmo de *Machine Learning* utilizado nesse trabalho para compor análise comparativa é o *Random Forest*, o qual se baseia em árvores de decisão e utiliza a técnica de *ensemble* para aumentar a acuracidade e reduzir o *overfitting*, ou seja, comportamento que gera baixa capacidade de generalização para dados fora do conjunto de treino.

Outra característica interessante desse modelo é seu aspecto randômico, durante a construção de cada uma das árvores de decisão, o algoritmo seleciona de forma aleatória um subconjunto de recursos em cada uma das ramificações, incrementando a robustez do modelo.

Dentre os parâmetros mais relevantes para o *Random Forest* temos o “*n_estimators*”, responsável por definir o número de árvores na floresta. Sendo válido contextualizar que um

número maior de árvores aumenta a precisão do modelo, porém também aumenta o custo de processamento.

De forma complementar, existe o parâmetro “*max_depth*”, que determina a profundidade máxima de cada árvore, controlando assim a complexidade das árvores e prevenindo o *overfitting*, além do “*max_features*”, responsável por determinar o número máximo de recursos considerados ao procurar a melhor divisão, introduzindo aleatoriedade e aumentando a diversidade entre as árvores, tornando o modelo mais robusto.

5. ANÁLISE COMPARATIVA

O foco desse capítulo se dará na implementação dos algoritmos de *Machine Learning* para previsão dos preços das ações da Petrobras e da Amazon, utilizando diferentes combinações de variáveis de entrada para cada um dos algoritmos apresentados na Seção 4.

O objetivo desse capítulo é entender como cada um dos algoritmos selecionados performa de acordo com a combinação de variáveis de entrada e identificar se há diferenças no comportamento das previsões para empresas de diferentes setores da economia.

Para a implementação dos algoritmos foi desenvolvido um código em linguagem de programação *Python*, que recebeu como entrada uma base de dados para cada uma das empresas analisadas com os preços de fechamento diário, os indicadores técnicos e os indicadores fundamentalistas conforme exposto na Seção 4.

Na implementação foram utilizadas as seguintes bibliotecas da linguagem de programação *Python*:

- “*pandas*”: para manipulação e análise de dados;
- “*yfinance*”: para extração dos dados históricos dos preços de fechamento e volume de negociação;
- “*ta*”: para cálculo dos indicadores técnicos;
- “*sklearn*”): para implementação dos algoritmos de *Machine Learning* e ferramentas relacionadas, mais especificamente as seguintes sub-bibliotecas:
 - “*train_test-split*”: para dividir os dados em subconjuntos de treino e teste;
 - “*GridSearchCV*”: ferramenta para realizar a busca em grande de parâmetros para otimização dos modelos;
 - “*RandomForestRegressor*”: para implementação do algoritmo *Random Forest* para regressão;
 - “*SVR*”: para implementação do algoritmo *SVR* para regressão;
 - “*StandardScaler*”: para normalização dos dados;
 - “*Pipeline*”: para encadeamento de múltiplas etapas de processamento e modelagem de dados.

No código desenvolvido foi utilizada técnica do *Pipeline* em conjunto com o *Grid Search* para otimização dos modelos. Enquanto a primeira é responsável, entre outras funções, por auxiliar no processo de normalização dos dados de entrada, a segunda é uma técnica que automatiza o processo de ajuste dos hiperparâmetros do modelo, usando validação cruzada para avaliar diferentes

combinações de hiperparâmetros. A combinação dessas duas técnicas torna o modelo mais robusto e traz uma maior acuracidade para as previsões.

Além disso, foram utilizados os seguintes valores para os hiperparâmetros:

- Para o algoritmo *SVR*:
 - “C” pode assumir os valores 0,1, 1 e 10, de acordo com o ajuste do *Grid Search* para cada uma das implementações;
 - “*epsilon*” assume o valor padrão da biblioteca de 0,1;
 - “*gamma*” assume o valor padrão da biblioteca de 1 sobre o número de variáveis de entrada.
- Para o algoritmo Random Forest:
 - “*n_estimators*” pode assumir os valores 100 e 200, de acordo com o ajuste do *Grid Search* para cada uma das implementações;
 - “*max_depth*” pode assumir os valores 2, 10 e 20, de acordo com o ajuste do *Grid Search* para cada uma das implementações;
 - “*max_features*” assume o valor padrão da biblioteca de número variáveis de entrada.

A análise comparativa será realizada por meio da implementação de cada um dos algoritmos com três diferentes combinações de variáveis de entrada, sendo elas:

- i) Apenas indicadores técnicos;
- ii) Apenas indicadores fundamentalistas;
- iii) Ambos.

Para permitir uma comparação assertiva entre a performance em cada um dos cenários de acordo com Parpinelli (2020), utilizou-se o coeficiente de determinação (R^2), uma métrica estatística que representa a proporção da variância para uma variável dependente que é explicada por uma ou mais variáveis independentes em um modelo de regressão. De forma simplificada, é uma métrica que indica a qualidade das variáveis de entrada em prever a variável dependente de um modelo de regressão. Sendo R^2 calculado pela equação:

$$R^2 = 1 - \frac{SS_r}{SS_t} \quad (8)$$

$$SS_r = \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (9)$$

$$SS_t = \sum_{i=1}^m (y_i - \bar{y})^2 \quad (10)$$

Onde:

y_i - são os valores reais

\hat{y}_i - são os valores previstos pelo modelo

\bar{y} - é a média dos valores reais

m - é a quantidade de valores previstos

SS_r - é a soma dos quadrados da diferença entre os valores reais e os valores previstos pelo modelo

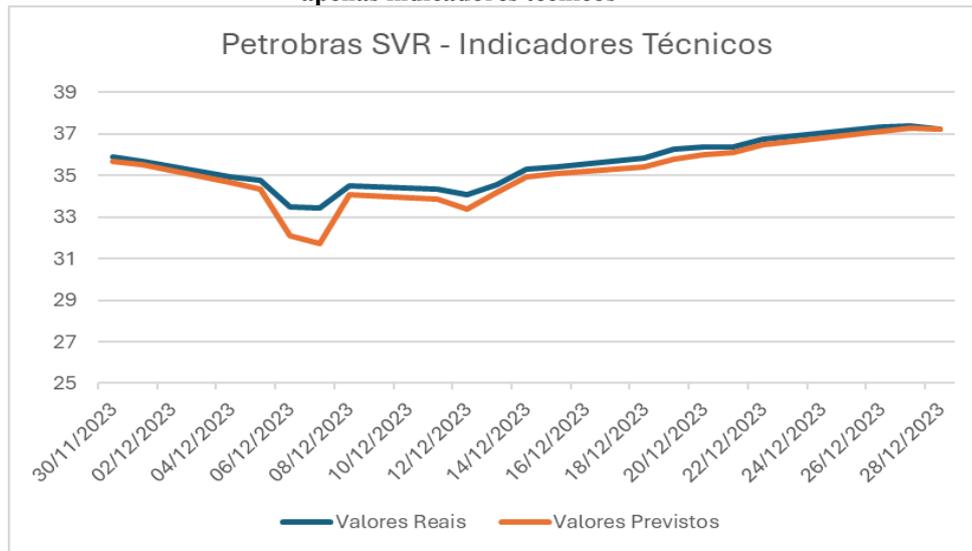
SS_t - é a soma dos quadrados da diferença entre os valores reais e a média dos valores observados

É fundamental contextualizar que, quanto mais próximo de 1 for o valor de R^2 , mais preciso na explicação da variabilidade dos dados é o modelo, ou seja, mais a linha de regressão se ajusta aos dados. No entanto, de acordo com Colin Cameron e Windmeijer (1997) os modelos podem apresentar valores fora do intervalo $[0,1]$ para R^2 em casos de regressão não linear, indicando uma performance pior do que uma simples média dos valores para casos em que R^2 é negativo, reflexo de um ajuste muito pobre do modelo em questão.

Além do coeficiente de determinação, são apresentados gráficos com os valores previstos e reais, para tornar a comparação mais visual e permitir uma maior clareza acerca da acuracidade das previsões em cada um dos cenários.

Inicialmente foi utilizado o algoritmo *SVR* recebendo como variável de entrada apenas os indicadores técnicos referentes a Petrobras, a amostra total utilizada compreende o período entre 02/01/2020 e 28/12/2023, enquanto a comparação entre os valores previstos e reais compreende o período entre 30/11/2023 e 28/12/2023.

Figura 4 – Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo SVR e apenas indicadores técnicos

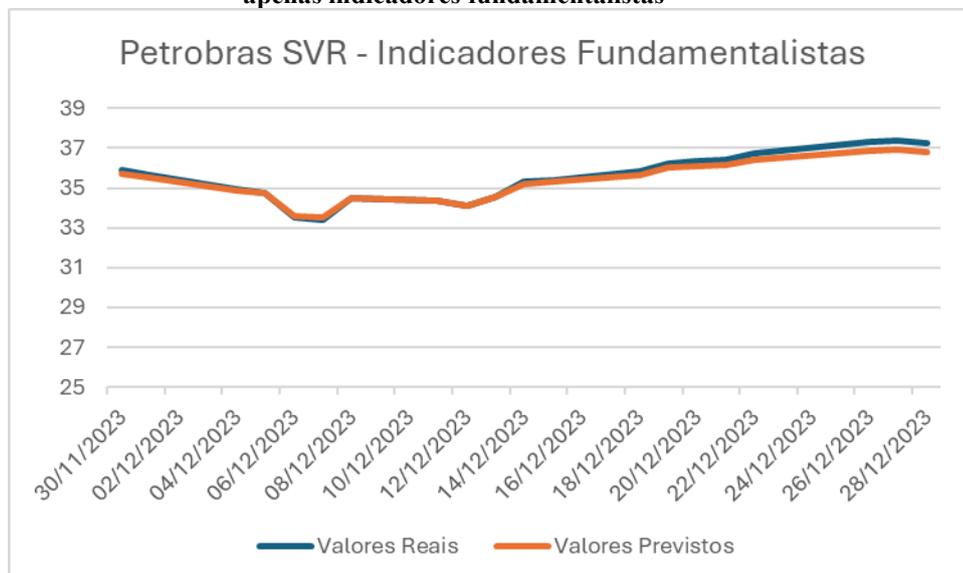


Fonte: Elaborado pelo autor (2024)

Neste primeiro comparativo representado pela Figura 4 é possível observar que o modelo foi capaz de incorporar bem a tendência de movimento do preço das ações, o que reflete a função de antecipação de tendências desempenhada pela maioria dos indicadores técnicos utilizados. Apesar do *gap* mais acentuado entre os dias 06/12/2023 e 08/12/2023, o modelo se manteve discretamente abaixo dos valores reais, refletindo uma previsão conservadora que pode ser bastante útil para investimentos de longo prazo. Além disso, o cenário apresentado resultou em um $R^2 = 0,8245$.

O segundo comparativo se deu por meio da implementação do algoritmo SVR utilizando apenas os indicadores fundamentalistas para o mesmo período com os dados da Petrobras.

Figura 5 - Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo SVR e apenas indicadores fundamentalistas

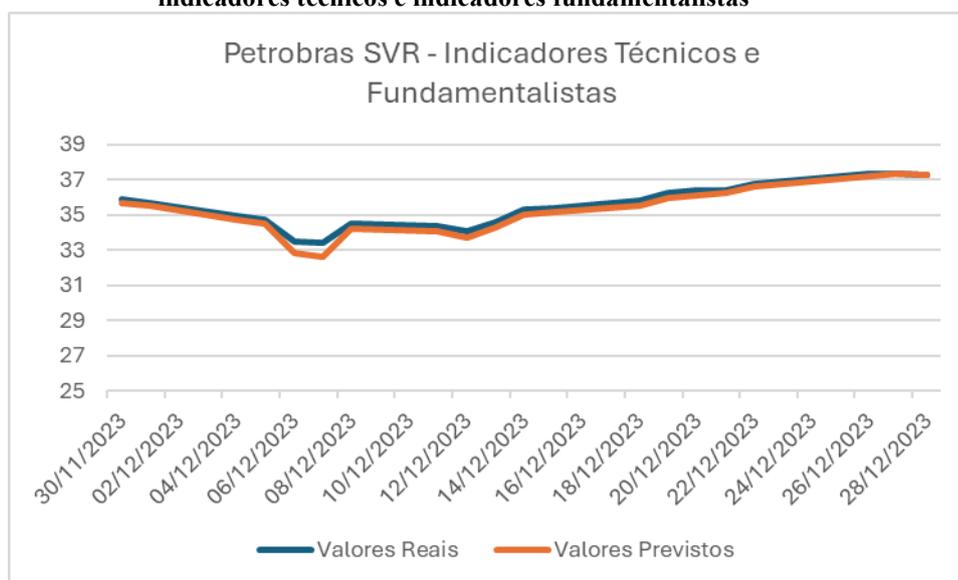


Fonte: Elaborado pelo autor

No comparativo ilustrado pela Figura 5 é possível observar um elevado nível de acuracidade das previsões, seguindo o acompanhamento da tendência do preço durante todo o período, com um pequeno *gap* nos últimos fechamentos. Estes resultados reforçam como os indicadores fundamentalistas podem gerar análises precisas quando aplicados no contexto de uma empresa de um setor mais tradicional, que apresenta níveis de crescimento mais estáveis. Nesta implementação foi obtido um $R^2 = 0,8028$.

No terceiro comparativo foi implementado o algoritmo *SVR* utilizando a combinação dos indicadores técnicos e fundamentalistas para o mesmo período com os dados da Petrobras.

Figura 6 – Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo *SVR*, indicadores técnicos e indicadores fundamentalistas



Fonte: Elaborado pelo autor

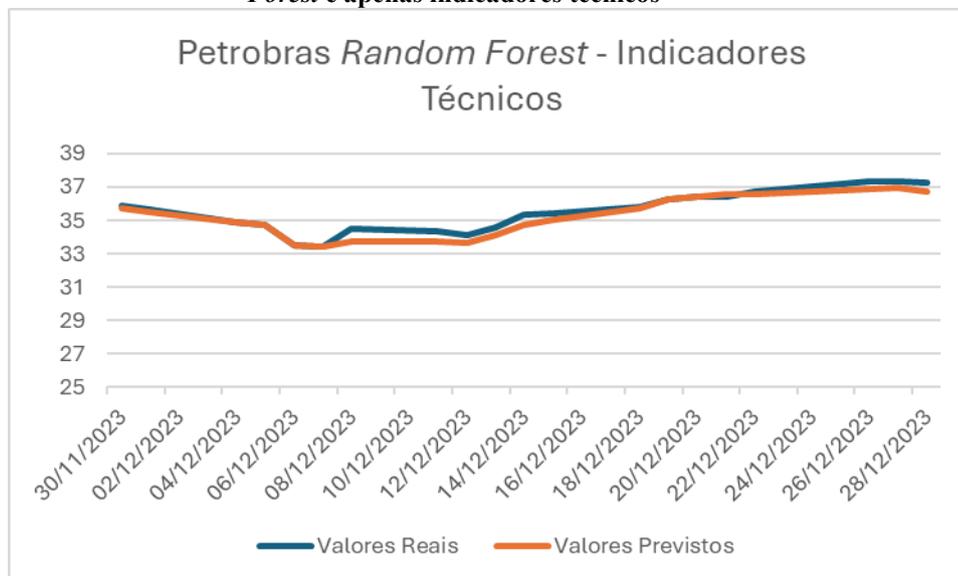
Neste último comparativo com a implementação do *SVR* utilizando dados da Petrobras conforme visto na Figura 6, observa-se um comportamento semelhante ao da Figura 4, porém com uma redução no *gap* observado entre os dias 06/12/2023 e 08/12/2023, indicando que a combinação dos indicadores nesse contexto acabou incrementado a acuracidade do modelo.

Foi obtido nesse cenário um $R^2 = 0,8336$, acima do apresentado nas duas implementações anteriores, indicando que a combinação dos indicadores resultou em um modelo com maior precisão na explicação da variabilidade dos dados.

Posteriormente, para complementação da análise comparativa, foi implementado o algoritmo *Random Forest* para o mesmo conjunto de dados da Petrobras com diferentes combinações de

variáveis de entrada. Sendo a primeira delas a implementação apenas utilizando os indicadores técnicos.

Figura 7 – Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo *Random Forest* e apenas indicadores técnicos

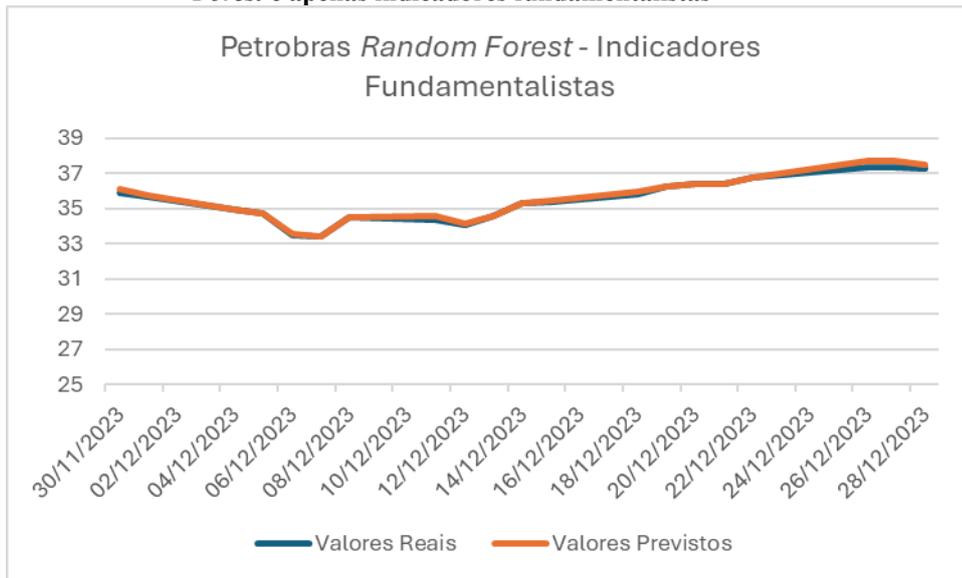


Fonte: Elaborado pelo autor

Analisando a Figura 7, nota-se que o modelo apresentou previsões bastante assertivas no início do período e mesmo com o *gap* apresentado entre os dias 08/12/2023 e 14/12/2023, seguiu acompanhando a tendência do preço ao real ao longo de todo o período. Nesta implementação foi obtido um $R^2 = 0,9104$, consideravelmente acima dos valores obtidos a partir do modelo *SVR*, indicando uma melhor performance de regressão do *Random Forest* nessa aplicação específica.

Dando continuidade a análise comparativa, há a implementação do *Random Forest* utilizando apenas os indicadores fundamentalistas referentes a Petrobras no mesmo período.

Figura 8 - Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo *Random Forest* e apenas indicadores fundamentalistas

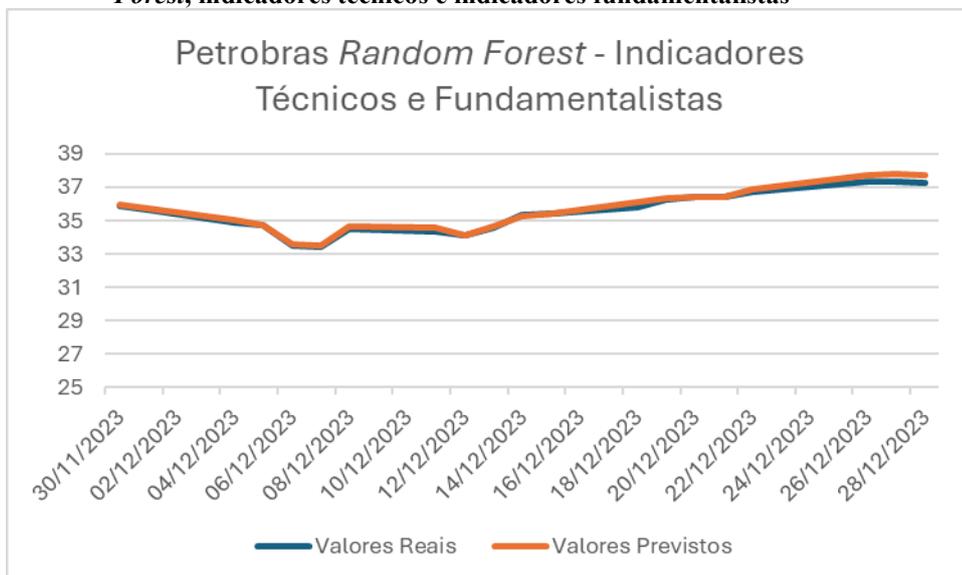


Fonte: Elaborado pelo autor

Na Figura 8 é possível perceber a assertividade do modelo ao longo de todo o período, mais uma vez reforçando a acuracidade dos indicadores fundamentalistas no contexto da Petrobras. Foi calculado nesta implementação um $R^2 = 0,9388$, reforçando a melhor performance do *Random Forest* quando comparados os resultados obtidos pelo *SVR* com esse mesmo conjunto de dados.

Para finalizar a análise comparativa utilizando os dados da Petrobras, foi realizada a implementação do *Random Forest* tendo indicadores técnicos e fundamentalistas como variáveis de entrada.

Figura 9 - Gráfico comparativo entre os valores reais e previstos para Petrobras utilizando o algoritmo *Random Forest*, indicadores técnicos e indicadores fundamentalistas



modelo em identificar as tendências de crescimento e decréscimo do preço da ação de maneira quase que imediata.

Foi obtido um $R^2 = 0,8326$ para esse cenário, bastante próximo do valor apresentado para esse mesmo cenário utilizando a amostra de dados da Petrobras, sendo um forte indicativo de que a utilização do *SVR* tendo como variáveis de entrada indicadores técnicos possui um comportamento semelhante independente do setor de atuação da empresa.

Dando seguimento à análise comparativa, será implementado o modelo *SVR* recebendo com variável de entrada apenas os indicadores fundamentalistas da amostra de dados da Amazon.

Figura 11 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo *SVR* e apenas indicadores fundamentalistas



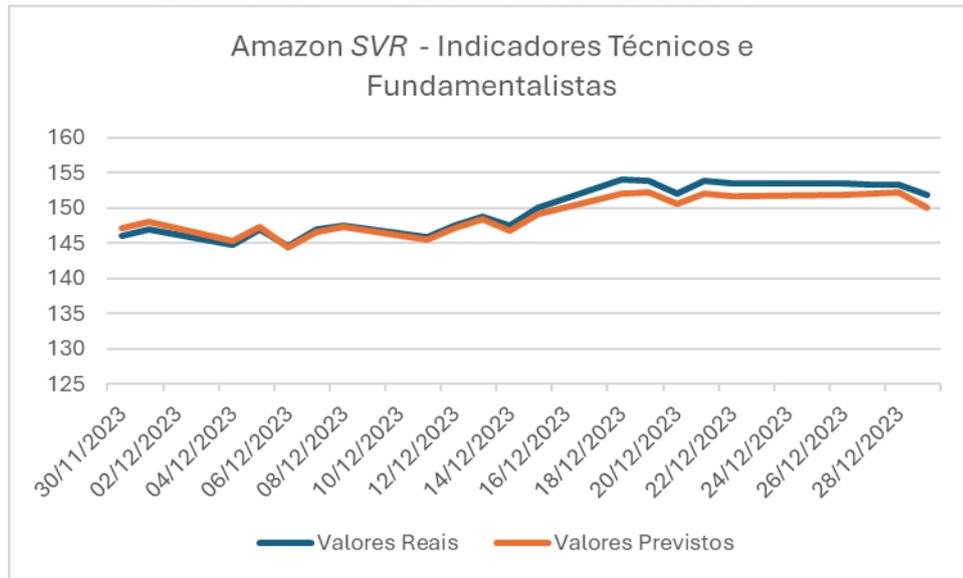
Fonte: Elaborado pelo autor

No cenário ilustrado pela Figura 11, é possível observar o comportamento constante dos valores previstos ao longo de todo o período comparativo, indicando que os indicadores fundamentalistas no contexto da Amazon foram incapazes de estabelecer qualquer tipo de conexão com a previsão dos preços de fechamento da ação, apenas introduzindo redundância e ruído ao modelo, sem contribuição para a capacidade preditiva do modelo.

O R^2 obtido nesse cenário foi igual a $-2,4888$, representando o ajuste pobre do modelo em questão, reforçando que a utilização de indicadores fundamentalistas mais tradicionais para análise de empresas do segmento de tecnologia é bastante ineficiente. Sendo crucial a utilização de indicadores mais sofisticados para o desenvolvimento de análises fundamentalistas neste setor.

Em seguida, será realizada a última implementação do algoritmo *SVR* utilizando indicadores técnicos e fundamentalistas como variáveis de entrada do modelo para amostra de dados da Amazon.

Figura 12 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo *SVR*, indicadores técnicos e indicadores fundamentalistas

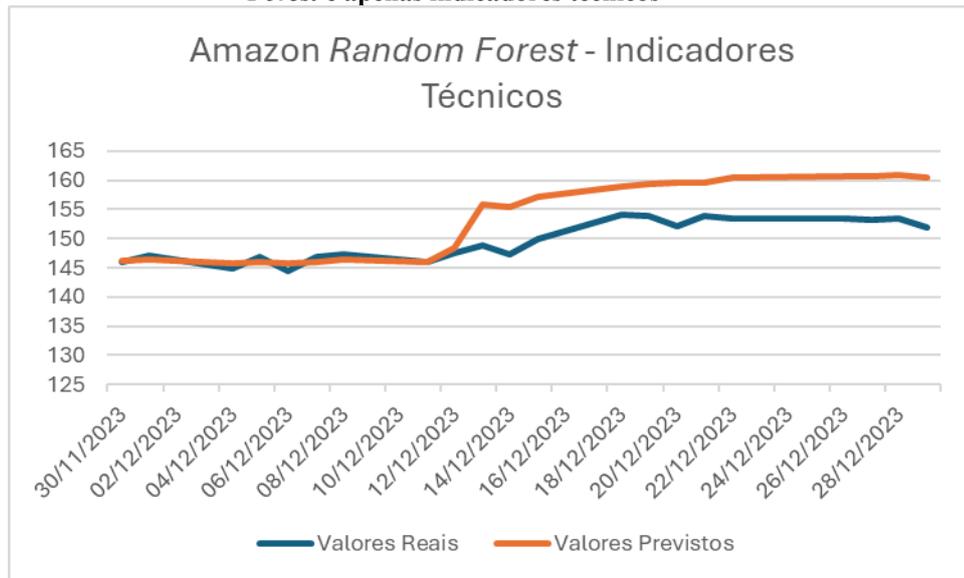


Fonte: Elaborado pelo autor

Conforme ilustrado na Figura 12, é bastante evidente como a combinação de indicadores técnicos e fundamentalistas resultou em um previsões bastante assertivas, indicando que o modelo os indicadores técnicos estão exercendo um papel dominante no modelo. Neste cenário foi obtido um $R^2 = 0,6124$, bem abaixo do resultado obtido com a implementação do *SVR* utilizando apenas indicadores técnicos, e indicando que a inclusão dos indicadores fundamentalistas reduziu a capacidade de ajuste da regressão aos dados.

Para um aprofundamento da análise comparativa, é fundamental realizar a implementação do algoritmo *Random Forest* utilizando as diferentes combinações de variáveis de entrada para a amostra de dados da Amazon. Sendo inicialmente implementado utilizando apenas os indicadores técnicos.

Figura 13 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo *Random Forest* e apenas indicadores técnicos



Fonte: Elaborado pelo autor

A partir da análise da Figura 13, é possível perceber *gaps* significativos ao longo do período, apesar do seguimento da tendência dos preços reais. Apesar de ter sido obtido um $R^2 = 0,9532$ para esse cenário, esse resultado pode representar um *overfitting*, ou seja, o modelo é capaz de se ajustar muito bem aos dados de treinamento, porém não é capaz de realizar previsões precisas utilizando novos dados.

A implementação seguinte se dará por meio da implementação do algoritmo *Random Forest*, recebendo como variável de entrada apenas os indicadores fundamentalistas.

Figura 14 – Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo *Random Forest* e apenas indicadores fundamentalistas

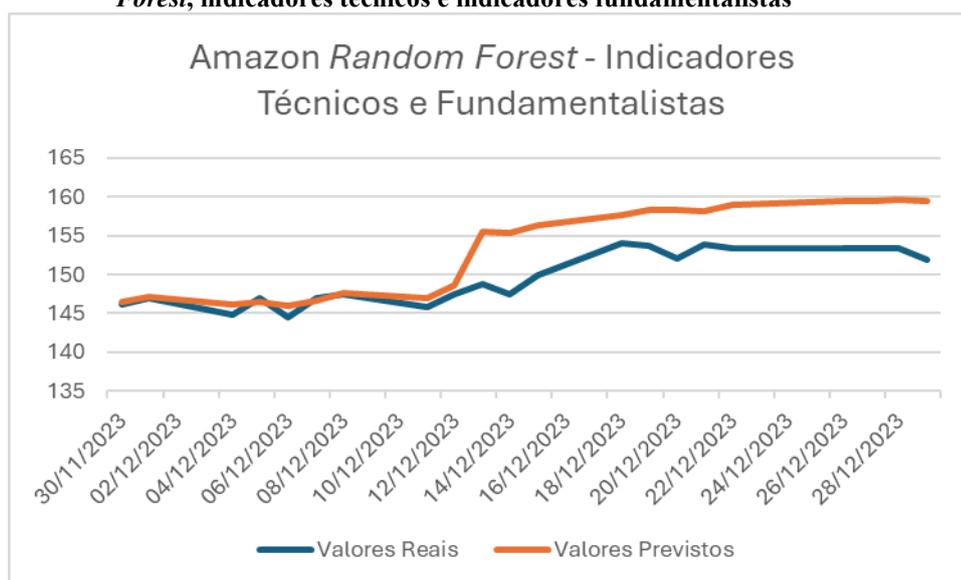


Fonte: Elaborado pelo autor

No cenário ilustrado pela Figura 14, é bastante interessante observar o elevado nível de acuracidade do modelo no início do período e posteriormente a repetição de valores constantes para as previsões do meio até o final do período. Sendo um forte indicativo de que os indicadores fundamentalistas do meio para o final do mês apresentaram um comportamento constante e tiveram uma redução na correlação com o preço da ação. Nesse cenário específico foi obtido um $R^2 = -0,0938$, reforçando a hipótese de que os indicadores fundamentalistas tradicionais não são eficientes na previsão dos preços das ações de empresas do setor de tecnologia.

Por fim, há a implementação do *Random Forest* para a amostra de dados da Amazon, utilizando a combinação de indicadores técnicos e fundamentalistas como variáveis de entrada do modelo.

Figura 15 - Gráfico comparativo entre os valores reais e previstos para Amazon utilizando o algoritmo *Random Forest*, indicadores técnicos e indicadores fundamentalistas



Fonte: Elaborado pelo autor

Nesse último comparativo, ilustrado pela Figura 15 é possível observar um comportamento bastante parecido com o apresentado na Figura 13, reforçando a predominância dos indicadores técnicos no modelo, uma vez que os indicadores fundamentalistas não foram eficientes no estabelecimento da sua correlação com os preços reais. Foi obtido nesta última implementação um $R^2 = 0,9429$, dando fortes indícios de *overfitting*, uma vez que as previsões acabaram não sendo tão assertivas.

Por fim, é possível elucidar os resultados obtidos em cada uma das implementações e compará-los de forma direta através dos valores de R^2 de forma agrupada na Tabela 2.

Tabela 2 – Resultado da análise comparativa

		Petrobras	Amazon
Algoritmo	Variável de Entrada	R ²	R ²
<i>SVR</i>	Indicadores Técnicos	0,8245	0,8326
	Indicadores Fundamentalista	0,8028	-2,4888
	Ambos	0,8336	0,6124
<i>Random Forest</i>	Indicadores Técnicos	0,9104	0,9532
	Indicadores Fundamentalista	0,9388	-0,0938
	Ambos	0,9380	0,9429

Fonte: Elaborado pelo autor (2024)

6. CONCLUSÃO

Ao longo deste trabalho, foi possível explorar a aplicação de modelos de Machine Learning, mais especificamente os algoritmos *Random Forest* e o *Support Vector Regression (SVR)*, com o objetivo de prever os preços das ações usando indicadores técnicos e fundamentalistas como variáveis de entrada.

Além disso, foi possível desenvolver um conhecimento mais aprofundado acerca da estrutura do mercado financeiro brasileiro em conjunto com o entendimento dos métodos de análise de empresas sob um viés técnico e fundamentalista.

Durante o desenvolvimento da análise comparativa, tornou-se claro que os modelos se comportam de forma bastante diferente, e sofrem uma forte influência do setor da empresa analisada, impactando a performance dos modelos.

De forma complementar, foi possível notar que a utilização de indicadores fundamentalistas tradicionais se mostrou bastante eficaz para ambos os algoritmos quando utilizados para previsão de preços de uma empresa de crescimento mais estável. No entanto, esses mesmos indicadores se mostraram incapazes de estabelecer uma correlação com os preços das ações para uma empresa de tecnologia, inserida em um cenário bem mais dinâmico e com múltiplos bem mais agressivos. Apesar disso, é de suma importância, para reforçar essas hipóteses, a realização de trabalhos futuros com foco na aplicação destes algoritmos em um maior número de empresas dos respectivos setores.

Enquanto os indicadores técnicos, se mostraram mais eficientes na captura da dinâmica de curto prazo do mercado, principalmente na identificação da tendência de alta e baixa dos preços.

Outro ponto bastante relevante que deve ser destacado, é a incidência de *overfitting* nas implementações do *Random Forest*, sendo uma sugestão de desenvolvimento futuro a aplicação de técnicas para reduzir esse comportamento e aprimorar a performance do modelo.

Algumas das limitações encontradas ao longo do desenvolvimento do trabalho foi a extração manual dos indicadores fundamentalistas, exigindo bastante tempo para o levantamento do histórico dos indicadores para cada uma das empresas analisadas, uma vez que as *APIs* existentes fornecem apenas os indicadores atuais, o que acabou por limitar o número de empresas analisadas em conjunto com a extensão da análise comparativa, a qual resultou em seis diferentes cenários por empresa.

Como trabalho futuro, sugere-se a replicação da implementação dos algoritmos para uma gama de empresas de um mesmo setor para validação do comportamento dos algoritmos em cada um dos setores analisados.

REFERÊNCIAS

AWAD, M.; KHANNA, R. **Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers**. Berkeley, Ca: Apressopen, 2015.

BEYAZ, E. *et al.* **Comparing Technical and Fundamental Indicators in Stock Price Forecasting**. 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), jun. 2018.

BRUNHUEMER, A. *et al.* **Supervised Machine Learning Classification for Short Straddles on the S&P500**. *Risks*, v. 10, n. 12, p. 235–235, 9 dez. 2022.

B3. **Guia Por dentro da B3**. Disponível em: <<https://www.b3.com.br/lumis/portal/file/fileDownload.jsp?fileId=8AE490CA7DB66DB8017DC9C044BB5C58>>. Acesso em: 6 mar. 2024.

B3. **Qual a estrutura do sistema financeiro nacional? Entenda!** Disponível em: <<https://borainvestir.b3.com.br/tipos-de-investimentos/estrutura-do-sistema-financeiro-nacional/>>. Acesso em: 6 mar. 2024.

CAUCHICK, Paulo Augusto *et al.* **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações**. Rio de Janeiro: Elsevier, 2011.

COLIN CAMERON, A.; WINDMEIJER, F. A. G. **An R-squared measure of goodness of fit for some common nonlinear regression models**. *Journal of Econometrics*, v. 77, n. 2, p. 329–342, 1 abr. 1997.

DAMODARAN, A. **Investment valuation: Tools and techniques for determining the value of any asset**. 3. ed. Hoboken, N.J.: Wiley, 2012.

DE OLIVEIRA, L. H. R.; RÊGO, T. G; DINIZ, J. A. **Previsão de Valores de Aquisições Governamentais: O Uso dos Conceitos de Data Science e Machine Learning**. XIX USP International Conference in Accounting, 2019.

DOMINGOS, P. **A few useful things to know about machine learning**. *Communications of the ACM*, v. 55, n. 10, p. 78, 1 out. 2012.

FABOZZI, F. J.; FOCARDI, S.; JONAS, C. **Trends in quantitative equity management: survey results**. *Quantitative Finance*, v. 7, n. 2, p. 115–122, abr. 2007.

GIL, A. C. **Como elaborar projetos de pesquisa**. 2ª ed., São Paulo: Atlas, 2002

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6ª ed., São Paulo: Atlas, 2009

R. HARIKRISHNAN et al. **Machine Learning Based Model to Predict Stock Prices: A Survey**. IOP conference series, v. 1084, n. 1, p. 012019–012019, 1 mar. 2021.

JAIN, P.; RAHMAN, I.; KULKARNI, B. D. **Development of a Soft Sensor for a Batch Distillation Column Using Support Vector Regression Techniques**. Chemical Engineering Research and Design, v. 85, n. 2, p. 283–287, jan. 2007.

LAKATOS, E. M.; ANDRADE, M. **Fundamentos de metodologia científica**. São Paulo (Sp): Editora Atlas S.A, 2010.

KUMBURE, M. M. et al. **Machine learning techniques and data for stock market forecasting: A literature review**. Expert Systems with Applications, v. 197, p. 116659, jul. 2022.

LEMOS, F. **Análise Técnica dos Mercados Financeiros: Um guia completo e definitivo dos métodos de negociação de ativos**. 2^a ed, São Paulo: Saraiva Educação, 2018.

MALTA, T. L.; DE CAMARGOS, M. A. **Variáveis da análise fundamentalista e dinâmica e o retorno acionário de empresas brasileiras entre 2007 e 2014**. REGE - Revista de Gestão, v. 23, n. 1, p. 52–62, jan. 2016.

PARPINELLI, F. **Dissecando modelos de Regressão Linear - parte 2**. Disponível em: <<https://medium.com/@felipeparpinelli/dissecando-modelos-de-regressão-linear-parte-2-e035b53dd351>>. Acesso em: 6 mar. 2024.

POLAMURI, S.; SRINIVAS, K.; MOHAN, A. **Stock market prices prediction using random forest and extra tree regression**. Int. J. Recent Technol. Eng, v. 8, n. 1, p. 1224-1228, 2019.

VIJH, M. et al. **Stock Closing Price Prediction using Machine Learning Techniques**. Procedia Computer Science, v. 167, n. 167, p. 599–606, 2020.

SANTOS, G. **Algoritmos de Machine Learning para previsão da B3**. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Uberlândia, Uberlândia, 2020.