



UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E
CONTABILIDADE
CURSO DE GRADUAÇÃO EM CIÊNCIAS ECONÔMICAS

MARCOS PAULO PINTO CHAVES

EFETIVIDADE DE MODELOS DE PREVISÃO EM FUNDOS IMOBILIÁRIOS:
UM ESTUDO COMPARATIVO ENTRE LSTM E ARIMA

FORTALEZA

2025

MARCOS PAULO PINTO CHAVES

EFETIVIDADE DE MODELOS DE PREVISÃO EM FUNDOS IMOBILIÁRIOS: UM
ESTUDO COMPARATIVO ENTRE LSTM E ARIMA

Trabalho de conclusão de curso apresentado ao Curso de Graduação em Ciências Econômicas da Universidade Federal do Ceará, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciências Econômicas.

Orientador: Prof. Victor Aguiar Evangelista de Farias

FORTALEZA

2025

Dados Internacionais de Catalogação na Publicação

Universidade Federal do Ceará

Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados
fornecidos pelo(a) autor(a)

C439e Chaves, Marcos Paulo Pinto.

Efetividade de modelos de previsão em fundos imobiliários: Um estudo
comparativo entre LSTM e ARIMA. / Marcos Paulo Pinto Chaves. – 2025.

46 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do
Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade,
Curso de Ciências Econômicas, Fortaleza, 2025. Orientação: Prof. Dr.
Victor Aguiar Evangelista de Farias.

1. Mercado Financeiro. 2. Fundos de Investimento Imobiliário. 3. Previsão.
4. LSTM. 5. ARIMA. I. Título.

CDD 330

MARCOS PAULO PINTO CHAVES

EFETIVIDADE DE MODELOS DE PREVISÃO EM FUNDOS IMOBILIÁRIOS: UM
ESTUDO COMPARATIVO ENTRE LSTM E ARIMA

Trabalho de conclusão de curso
apresentado ao Curso de Graduação
em Ciências Econômicas da
Universidade Federal do Ceará,
como parte dos requisitos
necessários para obtenção do grau de
Bacharel em Ciências Econômicas.

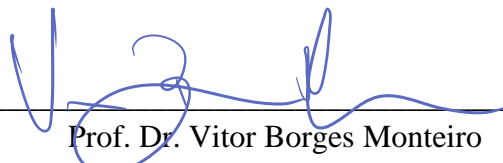
Orientador: Prof. Victor Aguiar
Evangelista de Farias

Aprovada em 21/02/2025.

BANCA EXAMINADORA

Prof. Dr. Victor Aguiar Evangelista de Farias (Orientador)

Universidade Federal do Ceará (UFC)



Prof. Dr. Vitor Borges Monteiro
Universidade Federal do Ceará (UFC)



Prof. Dr. Carlos de Oliveira Caminha Neto
Universidade Federal do Ceará (UFC)

AGRADECIMENTOS

Em primeiro lugar, gostaria de expressar minha profunda gratidão à minha mãe, Ana Paula, por ter me dado à luz, me criado com tanto amor e por sempre lutar incansavelmente pelo meu futuro, independentemente das circunstâncias. Ao meu saudoso pai, Marcílio José, agradeço por ter me ensinado valores fundamentais, por me mostrar o que significa ser um homem e por sempre me incentivar a trilhar o caminho da educação.

À minha irmã, Anelise, sou grato pelo apoio constante ao longo da elaboração deste trabalho, contribuindo de forma essencial com orientações sobre a escrita e a estrutura. À minha namorada, Maria Fernanda, agradeço por ter estado ao meu lado desde o início da faculdade, me incentivando a nunca desistir e a confiar em meu próprio potencial, mesmo nos momentos mais desafiadores. À Deus, agradeço por permitir toda esta trajetória de crescimento pessoal, acadêmico e profissional.

Por fim, agradeço ao meu orientador, Victor Aguiar, pelo apoio e orientação durante todas as etapas deste trabalho, desde a definição do tema até os ajustes finais, contribuindo de maneira fundamental para a realização deste TCC.

RESUMO

O mercado financeiro desempenha um papel fundamental na economia global ao facilitar a alocação eficiente de recursos. Ele serve como um mecanismo para conectar investidores com oportunidades de investimento, promovendo o crescimento econômico e a inovação. Dentro desse contexto, os fundos imobiliários são particularmente importantes, pois permitem que pequenos investidores acessem o mercado imobiliário com maior facilidade e diversificação. Portanto, prever os retornos dos FII's pode ajudar o investidor a selecionar os ativos a serem selecionados para sua carteira de investimentos. Este estudo compara a efetividade de diferentes modelos de previsão de preços em fundos imobiliários, avaliando a capacidade preditiva para os maiores fundos por patrimônio líquido de cada setor de atuação, conforme definidos pela ANBIMA. Para isso, foram utilizados os modelos *Long Short-Term Memory (LSTM)* e *AutoRegressive Integrated Moving Average (ARIMA)* para prever o preço de fechamento do dia seguinte. Modelos foram desenvolvidos para cada um desses fundos utilizando diferentes janelas deslizantes, com o objetivo de identificar quais abordagens oferecem os melhores resultados preditivos. A métrica de erro empregada na comparação dos modelos foi o erro quadrático médio (MSE). Os resultados da análise indicam que o modelo ARIMA supera consistentemente o modelo LSTM em todas as comparações, demonstrando uma maior precisão na previsão dos preços de fechamento dos fundos imobiliários analisados.

Palavras-chave: Mercado Financeiro, Fundos de Investimento Imobiliário, Previsão, LSTM, ARIMA, MSE.

ABSTRACT

The financial market plays a fundamental role in the global economy by facilitating the efficient allocation of resources. It serves as a mechanism to connect investors with investment opportunities, promoting economic growth and innovation. Within this context, real estate investment funds are particularly important as they allow small investors to access the real estate market with greater ease and diversification. In this scenario, predicting the returns of Brazilian real estate investment funds - *Fundo de Investimento Imobiliário (FII)*, can help investors select the assets to be included in their investment portfolios. This study compares the effectiveness of different price prediction models in real estate investment funds, evaluating the predictive capacity for the largest equity funds in each sector, as defined by ANBIMA. For this purpose, Long Short-Term Memory (LSTM) and AutoRegressive Integrated Moving Average (ARIMA) models were used to predict the next day's closing price (D+1). Models were developed for each of these funds using different sliding windows, aiming to identify which approaches offer the best predictive results. The error metrics used in the comparison of the models were the mean squared error (MSE). The analysis results indicate that the ARIMA model consistently outperforms the LSTM model in all comparisons, demonstrating greater accuracy in predicting the closing prices of the analyzed real estate investment funds.

Keywords: Financial Market, Real Estate Investment Funds, Forecasting, LSTM, ARIMA, MSE.

LISTA DE FIGURAS

Figura 1 - Modelo matemático de um neurônio	17
Figura 2 - Redes neurais recorrentes	18
Figura 3- Redes <i>Long Short-Term Memory</i>	19
Figura 4 - 1ª Etapa do processo de k-fold cross-validation	23
Figura 5 - 2ª Etapa do processo de k-fold cross-validation	24
Figura 6 - 3ª Etapa do processo de k-fold cross-validation	24
Figura 7 - 4ª Etapa do processo de k-fold cross-validation	24
Figura 8 - 5ª Etapa do processo de k-fold cross-validation	25
Figura 9 - Série completa de preços de fechamento do KNIP11.....	30
Figura 10 - Série completa de preços de fechamento normalizados do KNIP11	31
Figura 11 - Função de autocorrelação das séries temporais analisadas.....	42

LISTA DE TABELAS

Tabela 1 - Informações sobre os fundos que compõem o IFIX.....	27
Tabela 2 - Dados completos dos fundos que compõem o IFIX.....	28
Tabela 3 - Informações sobre os fundos escolhidos	29
Tabela 4 - Exemplo de histórico de preços do fundo KNIP11.....	29
Tabela 5 - Hiperparâmetros LSTM Testados	31
Tabela 6 Hiperparâmetros ARIMA Testados	32
Tabela 7 - Erros medidos pelos modelos LSTM e ARIMA para o fundo HFOF11.SA	33
Tabela 8 - Melhores modelos LSTM – HFOF11	34
Tabela 9 - Melhores modelos ARIMA – HFOF11	34
Tabela 10 - Erros medidos pelo melhor modelo LSTM e ARIMA para o fundo HGLG11.SA.....	35
Tabela 11 - Melhores modelos LSTM – HGLG11	35
Tabela 12 - Melhores modelos ARIMA - HGLG11	36
Tabela 13 - Erros medidos pelo melhor modelo LSTM e ARIMA para o fundo KNIP11.SA.....	36
Tabela 14 - Melhores modelos LSTM - KNIP11	37
Tabela 15 - Melhores modelos ARIMA - KNIP11	37
Tabela 16 - Erros medidos pelo melhor modelo LSTM e ARIMA para o fundo KNRI11.SA	38
Tabela 17 - Melhores modelos LSTM - KNRI11	38
Tabela 18 - Melhores modelos ARIMA - KNRI11	39
Tabela 19 - Erros medidos Erros medidos pelo melhor modelo LSTM e ARIMA para o fundo MFII11.SA	39
Tabela 20 - Melhores modelos LSTM - MFII11	40
Tabela 21 - Melhores modelos ARIMA - MFII11	40

SUMÁRIO

1 INTRODUÇÃO	11
1.1 Motivação para o Trabalho	11
1.2 Objeto de Estudo	12
1.3 Objetivos	12
1.4 Hipóteses	12
1.5 Organização do Trabalho	12
2 FUNDAMENTAÇÃO TEÓRIA	13
2.1 Mercado Financeiro	13
2.1.1 Fundos de Investimento Imobiliário	13
2.1.2 Eficiência do Mercado Financeiro	14
2.2 Python	15
2.3. Séries Temporais	15
2.4 Janelas Deslizantes	16
2.5 Redes Neurais Artificiais	16
2.5.1 Redes Neurais Recorrentes	17
2.5.2 LSTM	18
2.6 ARIMA	21
2.7 Validação dos modelos	22
2.8 Discussão Teórica sobre previsão de preços no Mercado Financeiro	25
3 METODOLOGIA	27
3.1 Base de Dados	27
3.2 Tratamento dos dados	30
3.3 Construção dos modelos	31
3.3.1 Modelo LSTM	31
3.3.2 Modelo ARIMA	32
4 AVALIAÇÃO E DISCUSSÃO DOS RESULTADOS	33
4.1 Resultados obtidos com o fundo HFOF11	33
4.2 Resultados obtidos com o fundo HGLG11	35
4.3 Resultados obtidos com o fundo KNIP11	36
4.4 Resultados obtidos com o fundo KNRI11	38
4.5 Resultados obtidos com o fundo MFII11	39
5 CONCLUSÃO	41
6 TRABALHOS FUTUROS	43
7 REFERÊNCIAS	45

1 INTRODUÇÃO

1.1 Motivação para o Trabalho

Fundos Imobiliários (FIIs) são um tipo de fundo de investimento em que os recursos dos investidores são aplicados no mercado imobiliário, através de compras de galpões, construções de edifícios, títulos de crédito imobiliário entre outros. A previsão dos preços desses ativos é uma tarefa desafiadora, dada a natureza volátil e não linear dos dados financeiros. “Os riscos do mercado financeiro referem-se aos fatores macroeconômicos, como eventos de natureza política, econômica e financeira que alteram a ordem atual, incluindo variações nas taxas de juros, modificações da legislação, entre outros” (Manganotti, 2014). Tendo em vista que os preços dos ativos podem não ser definidos de forma aleatória, é possível identificar uma oportunidade de prever os preços no mercado financeiro. Além disso, prever os retornos dos FII's pode ajudar o investidor a selecionar os ativos a serem selecionados para sua carteira (Cici et al, 2011).

Modelos tradicionais, como ARIMA tem sido amplamente utilizado para esse fim. O ARIMA é um modelo estatístico que prevê valores futuros com base em padrões históricos, assumindo que os dados seguem uma estrutura linear e estacionária, ou seja, possuem propriedades estatísticas constantes ao longo do tempo (Box et al., 1976).

Com o avanço das técnicas de aprendizado de máquina, surgiu o LSTM (*Long Short-Term Memory*), uma rede neural recorrente projetada para capturar padrões complexos em séries temporais. Diferente dos modelos tradicionais, o LSTM é capaz de identificar e aprender relacionamentos não lineares de longo prazo nos dados (Hochreiter; Schmidhuber, 1997). Diante dessas diferenças, torna-se necessário avaliar qual desses modelos oferece previsões mais acuradas, a fim de mensurar sua capacidade preditiva em relação aos preços de fundos imobiliários.

O modelo LSTM já foi utilizado para previsão de preços de fundos imobiliários, indicando um potencial das LSTMs para tomada de decisão diária e ganhos consistentes no mercado de fundos imobiliários brasileiro (Kondrat, 2021). Estudos recentes, como o de Rhanoui *et al.* (2020), que aplicaram o modelo LSTM na previsão de séries temporais financeiras, demonstram que essa abordagem é particularmente eficaz na captura de padrões temporais complexos, superando métodos tradicionais como o já citado, ARIMA, em termos de precisão preditiva.

Considerando esse contexto, tem-se a seguinte questão de pesquisa: Qual modelo de previsão LSTM e ARIMA oferece maior acurácia na previsão dos preços de

fechamento de fundos imobiliários no mercado brasileiro? Esse trabalho tem por finalidade esclarecer essa questão e mostrar a magnitude da melhoria da previsão e em quais situações cada modelo se prevalece.

1.2 Objeto de Estudo

O objeto de estudo para previsão será o preço de fechamento dos ativos em D+1 (dia útil posterior à data atual), conforme também foi utilizado por Amorim *et al.* (2023). Esses preços refletem a cotação da última negociação do ativo durante o dia e capturam um consenso do mercado do preço justo do ativo naquele dia, conforme definido por Amorim *et al.* (2023).

1.3 Objetivos

Sendo assim, objetivo geral do trabalho consiste em: Comparar a eficácia dos modelos LSTM, ARIMA na previsão dos preços de fechamento de fundos imobiliários, identificando qual modelo proporciona maior precisão e confiabilidade. Tem se como objetivos específicos:

- a) Avaliar o desempenho do modelo LSTM na previsão de preços de fundos imobiliários.
- b) Avaliar o desempenho do modelo ARIMA na previsão de preços de fundos imobiliários, verificando sua acurácia em relação a dados financeiros históricos.
- c) Entender sob quais condições cada modelo se destaca ou apresenta limitações.

1.4 Hipóteses

As hipóteses do estudo que serão investigadas são:

- a) O mercado financeiro não é realmente eficiente, e dessa forma, variações de preços de ativos podem sim ser previstas até certo nível, (Lo; MacKinlay, 1999).
- b) O modelo LSTM apresentará maior acurácia na previsão de preços de fundos imobiliários em comparação com o modelo ARIMA pela maior capacidade de capturar padrões complexos, como demonstrado por Nascimento *et al* (2022).

1.5 Organização do Trabalho

Este trabalho será estruturado em cinco seções principais: Fundamentação Teórica, Metodologia, Avaliação e Discussão dos Resultados, Conclusão e Trabalhos Futuros. Na seção de Fundamentação Teórica, serão explorados os conceitos e teorias

fundamentais que sustentam o tema do estudo. A seção de Metodologia descreverá os métodos e técnicas empregados na coleta e análise dos dados. A seção de Avaliação e Discussão dos Resultados tem o objetivo de apresentar os resultados obtidos por meio da metodologia aplicada e, com base nesses dados, analisar a capacidade preditiva dos modelos. A Conclusão proporcionará um resumo abrangente das análises e principais achados do estudo. Por fim, a seção de Trabalhos Futuros irá sugerir possíveis continuações e extensões do estudo.

2 FUNDAMENTAÇÃO TEÓRIA

Nesta seção, serão apresentados os conceitos essenciais para a compreensão deste trabalho, abordando os principais fundamentos dos fundos imobiliários, o funcionamento do mercado de fundos de investimento, bem como os modelos de previsão aplicados, incluindo o modelo LSTM e o modelo ARIMA. Além disso, serão exploradas as métricas de avaliação empregadas para medir a precisão e a efetividade das previsões, de modo a estabelecer uma base teórica robusta para a análise comparativa proposta entre os modelos de previsão.

2.1 Mercado Financeiro

Conforme Levin et al. (2002), o mercado financeiro desempenha um papel fundamental na economia global, sendo responsável pela intermediação de recursos entre agentes superavitários e deficitários. Ele é composto por um conjunto de instituições, instrumentos e operações que facilitam a negociação de ativos financeiros, como ações, títulos de dívida, moedas e derivativos. Segundo Mishkin (2019), o mercado financeiro é essencial para a alocação eficiente de recursos, uma vez que permite o financiamento de investimentos produtivos por meio da captação de poupança.

2.1.1 Fundos de Investimento Imobiliário

De acordo com a Instrução CVM (Órgão normativo e supervisor dos valores mobiliários) 472, de 31 de outubro de 2008, que estabelece as Normas gerais para regulamentação dos Fundos de Investimentos Imobiliários (FIIs), esses fundos são definidos como a comunhão de recursos captados por meio do sistema de distribuição de valores mobiliários, destinados à aplicação em empreendimentos imobiliários, ressaltando que 95% da renda mensal deve ser distribuída aos cotistas. Esses fundos podem ser classificados em 5 diferentes estratégias de investimento, fundos de tijolo,

fundos de papel (também chamados de fundos de títulos e valores mobiliários), fundos híbridos, fundos de fundos e fundos de desenvolvimento, conforme definido pela ANBIMA (Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais) em 2015.

Os fundos de tijolos são os fundos que fazem investimentos em imóveis físicos, como shoppings e galpões, gerando receita através da locação desses imóveis. Já os fundos de papel são os que investem em títulos e valores mobiliários atrelados ao setor imobiliário, como Certificados de Recebíveis Imobiliários (CRI) e Letras de Crédito Imobiliário (LCI). Os fundos de fundos são os que investem em cotas de outros tipos de fundos imobiliários, permitindo uma diversificação maior para os cotistas. Por fim, os fundos de desenvolvimento são aqueles que focam na construção e desenvolvimento de novos empreendimentos imobiliários, buscando altos retornos na entrega de projetos.

Ademais, os fundos de investimento imobiliários possuem um índice próprio (IFIX) na bolsa de valores, que contém uma carteira teórica de ativos, onde o objetivo desse índice é trazer um índice médio de retorno para os fundos imobiliários brasileiros.

2.1.2 Eficiência do Mercado Financeiro

A eficiência do mercado financeiro é uma questão amplamente debatida na literatura econômica. A Hipótese dos Mercados Eficientes (HME), proposta por Fama (1970), sugere que os preços dos ativos refletem todas as informações disponíveis, tornando impossível prever consistentemente movimentos de preços com base em informações passadas. De acordo com a HME, os preços dos ativos se ajustam rapidamente a qualquer nova informação, o que implica que nenhuma estratégia de análise técnica ou fundamental seria capaz de gerar retornos excessivos de forma sistemática.

No entanto, estudos recentes têm questionado essa hipótese. Machado e Correa (2022), por exemplo, conseguiram prever preços de ações com precisão significativa utilizando os modelos ARIMA e LSTM, indicando que algumas abordagens de previsão podem capturar padrões temporais relevantes nos preços dos ativos. Da mesma forma, Silveira (2021) explorou a viabilidade do uso de técnicas de *machine learning*, como Random Forest e LSTM, no mercado de ações brasileiro, ressaltando o potencial dessas metodologias em contextos financeiros complexos. Esses estudos demonstram que, mesmo em mercados teoricamente eficientes, metodologias avançadas podem oferecer vantagens preditivas.

2.2 Python

Entre as diversas ferramentas disponíveis para análise de dados e previsões, como a linguagem R, Eviews e Structured Query Language (SQL), optou-se pela linguagem de programação Python. Essa escolha se deve à vasta quantidade de trabalhos relacionados a *machine learning* que utilizaram Python como ferramenta principal, como evidenciado em estudos de Tavares e Quadrarelli (2022) e Moura (2023). O Python oferece uma ampla variedade de bibliotecas com capacidades preditivas, como TensorFlow e Keras (Abadi et al., 2015; Chollet, 2015), além do Statsmodels (Seabold; Perktold, 2010), que serão empregados para treinar os modelos LSTM e ARIMA, respectivamente. Além disso, por ser uma linguagem de código aberto, qualquer pessoa pode acessar e utilizar os códigos desenvolvidos neste trabalho. Os códigos desenvolvidos ao longo deste trabalho estão disponíveis publicamente no GitHub, uma plataforma de hospedagem de código-fonte baseada em repositórios. O acesso ao repositório pode ser feito através do link: https://github.com/sepectrok/codigos_tcc.

2.3. Séries Temporais

Conforme Wooldridge (2000), séries temporais são compostas por uma sequência de observações de um determinado fenômeno financeiro ao longo do tempo. Matematicamente, uma série temporal pode ser representada como uma sequência ordenada de valores:

$$X = X_1, X_2, \dots, X_t.$$

Onde X representa a série temporal, X_1 representa o valor da série no tempo 1, X_2 representa o valor da série no tempo 2 e X_t representa o valor da série no tempo t . A análise dessas séries temporais financeiras envolve a identificação de padrões e tendências nos dados, bem como a decomposição dos componentes que influenciam o comportamento desses dados. Segundo Hyndman e Athanaspoulos (2018), ao escolher um método de previsão de séries temporais, primeiro precisamos identificar os padrões de série temporal nos dados e, em seguida, escolher um método que seja capaz de capturar os padrões de maneira adequada. Portanto, ao analisar séries temporais e as possibilidades de previsão, é necessário utilizar modelos que sejam capazes de capturar esses padrões.

2.4 Janelas Deslizantes

A técnica de janelas deslizantes é amplamente utilizada em modelos de séries temporais, como LSTM e ARIMA, para preparar os dados de entrada de forma que seja possível capturar padrões históricos relevantes. Essa abordagem consiste em criar subconjuntos sequenciais de dados, onde cada janela contém um número fixo de observações passadas e a previsão corresponde ao valor imediatamente posterior à última observação da janela.

A implementação dessa técnica pode ser descrita em dois passos principais:

1. Definição das janelas: Para cada ponto t na série temporal, seleciona-se um conjunto de n observações anteriores $X_t = x_{t-1}, x_{t-2}, \dots, x_{t-n}$ como entrada, enquanto a saída correspondente y_t será o valor da série no ponto t .
2. Iteração sequencial: Esse processo é repetido ao longo de toda a série, criando pares (X_t, y_t) que são usados para treinar e avaliar os modelos.

Por exemplo, com uma janela de tamanho $t = 3$ em uma série temporal $[x_1, x_2, x_3, x_4, x_5]$, com tamanho 5 as janelas e saídas geradas seriam:

- Entrada: $X_4 = [x_3, x_2, x_1]$ Saída: y_4 .
- Entrada: $X_5 = [x_4, x_3, x_2]$ Saída: y_5 .

Assim como realizado no trabalho de Vasco (2020), diferentes tamanhos de janelas serão testados para comparar os resultados obtidos com os melhores modelos em cada série temporal. O uso de tamanhos variados, como 10, 30 e 60 observações, permite avaliar a influência da quantidade de informações históricas na capacidade preditiva dos modelos. Tamanhos menores podem capturar padrões de curto prazo, enquanto tamanhos maiores podem ser mais adequados para identificar tendências de longo prazo.

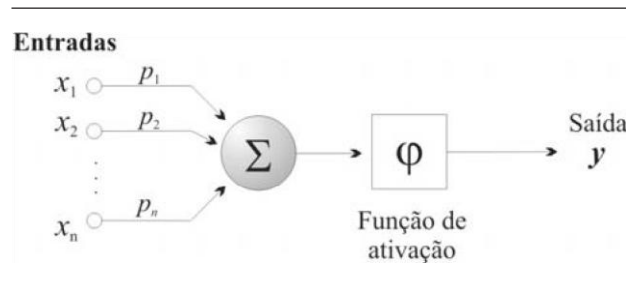
2.5 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados no funcionamento do cérebro humano, compostos por unidades interconectadas chamadas neurônios artificiais. Esses modelos são amplamente utilizados em problemas de reconhecimento de padrões, classificação e previsão, devido à sua capacidade de aprender e generalizar a partir de dados. É possível observar essas redes sendo utilizadas para previsão de falências, avaliação de crédito e até mesmo detecção de fraude (Santos *et al.* 2016).

As RNAs destacam-se em tarefas onde as relações entre as variáveis são não lineares e complexas, tornando-se uma ferramenta essencial em diversas áreas, como

visão computacional, processamento de linguagem natural e finanças (Haykin, 1998). Elas podem ser projetadas com diferentes arquiteturas, dependendo do problema a ser resolvido. Por exemplo, redes neurais *feedforward* possuem um fluxo de informações unidirecional, formando um grafo acíclico, como pode ser observado na Figura 1.

Figura 1 - Modelo matemático de um neurônio



Fonte: FERNEDA, Edberto, 2006

Essa imagem representa o funcionamento básico de um neurônio artificial em uma Rede Neural Artificial (RNA).

- **Entradas** ($x_1, x_2 \dots x_n$): São os dados fornecidos ao neurônio, cada um com seu respectivo peso ($p_1, p_2 \dots p_n$), que determina a importância de cada entrada.
- **Soma ponderada** (Σ): O neurônio calcula a soma ponderada das entradas e seus pesos.
- **Função de ativação** (ϕ): A soma passa por uma função de ativação, que aplica uma transformação matemática, introduzindo não linearidade ao modelo.
- **Saída** (y): O resultado da função de ativação é a saída do neurônio, usada como entrada para outros neurônios ou como resultado da rede.

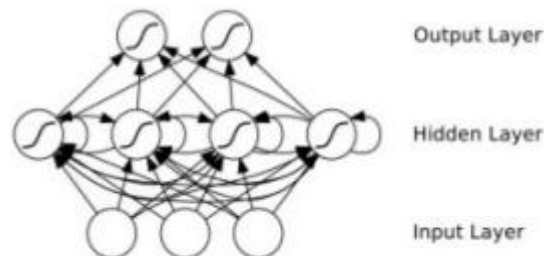
2.5.1 Redes Neurais Recorrentes

As Redes Neurais Recorrentes (RNNs) são uma classe específica de redes neurais artificiais projetadas para lidar com dados sequenciais, como séries temporais. Diferentemente das redes neurais tradicionais, onde entradas e saídas são tratadas de forma independente, as RNNs possuem conexões cíclicas em sua arquitetura, permitindo que informações sejam retidas e reutilizadas ao longo do tempo. Essa característica as torna particularmente eficazes em tarefas que dependem de contexto histórico, como a modelagem de séries temporais financeiras (Goodfellow et al., 2016).

Uma das principais vantagens das RNNs é a presença de uma memória interna que permite capturar dependências temporais de curto e longo prazo nos dados. Como

pode ser visto na figura 2, a informação não flui em um único sentido, e a saída da rede não depende mais apenas da entrada corrente, mas também das entradas anteriores, fornecendo uma interdependência temporal.

Figura 2 - Redes neurais recorrentes



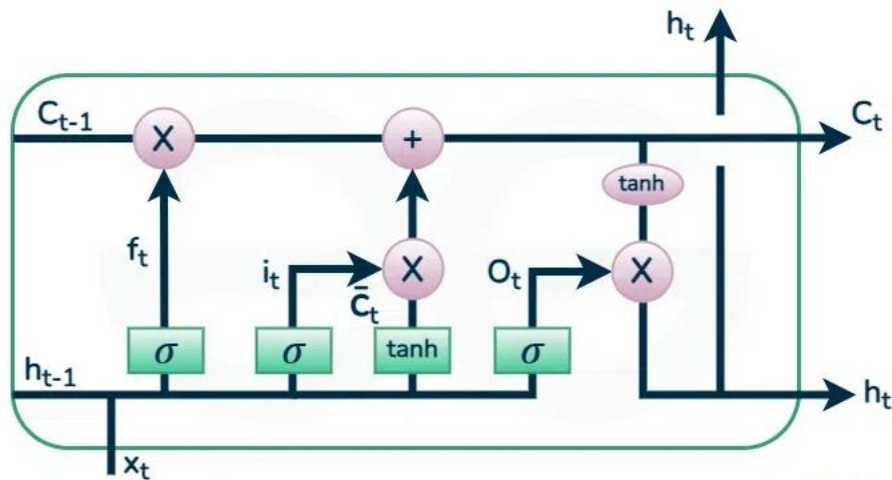
Fonte: NELSON, David, 2017.

Contudo, as RNNs tradicionais enfrentam limitações ao lidar com dependências de longo prazo, devido ao problema do gradiente desaparecido ou explodido durante o treinamento (Bengio et al., 1994). Para superar esses desafios, foram desenvolvidas variantes mais avançadas, como a *Long Short-Term Memory* (LSTM) e a *Gated Recurrent Unit* (GRU). Essas arquiteturas introduzem mecanismos de controle, como *gates* (portas), que ajudam a selecionar e preservar informações relevantes por períodos prolongados, tornando-as mais eficazes para a previsão de séries temporais complexas.

2.5.2 LSTM

A *Long Short-Term Memory* (LSTM) é uma arquitetura específica de Redes Neurais Recorrentes (RNNs) desenvolvida para superar as limitações das RNNs tradicionais, especialmente no que diz respeito ao aprendizado de dependências de longo prazo em dados sequenciais. Proposta por Hochreiter e Schmidhuber (1997), a LSTM introduz células de memória capazes de manter informações por longos períodos, sem que o problema do gradiente desaparecido ou explodido afete o processo de treinamento.

Figura 3- Redes *Long Short-Term Memory*



Fonte: GEEKSFORGEEEKS, 2024

A arquitetura LSTM é composta por uma série de "portas" – a porta de entrada, a porta de esquecimento e a porta de saída – que regulam o fluxo de informações dentro da célula de memória. A partir da arquitetura da rede LSTM apresentada na Figura 3, é demonstrado que essa estrutura funciona da seguinte forma:

Entradas:

- x_t : Representa a entrada no tempo atual.
- h_{t-1} : A saída da célula LSTM no instante anterior t-1, também chamada de estado oculto.
- C_{t-1} : O estado da célula em t-1, que transporta informações de longo prazo.

Gates (portas):

- **Porta de Esquecimento (f_t):** Decide quais informações do estado anterior (C_{t-1}) serão descartadas ou mantidas. É calculada da seguinte forma: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f$. Onde W_f e b_f são os pesos e o bias da porta de esquecimento. A ativação sigmoide σ retorna valores entre 0 e 1, indicando a proporção da informação a ser esquecida.
- **Porta de Entrada i_t :** Controla quais novas informações provenientes de x_t serão adicionadas ao estado atual. Consiste em dois componentes:
 - $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i$. Que determina a relevância da entrada, variando de 0 a 1.
 - $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C$. Que cria um novo candidato para atualização do estado da célula.

As informações relevantes são combinadas multiplicativamente ($i_t \cdot \tilde{C}_t$)

- **Porta de Saída o_t :** Determina quais informações do estado interno C_t serão usadas para calcular a saída h_t . É calculada como:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

Atualização da Célula C_t : A célula que absorve e atualiza as informações passadas e as utiliza para prever as posteriores é calculada da seguinte forma: $C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$
Saída h_t : É o valor estimado pelo modelo e é definido através da seguinte equação: $h_t = o_t \cdot \tanh(C_t)$.

Esse mecanismo complexo permite que a LSTM retenha e acesse informações relevantes ao longo de longas sequências de dados, tornando-a particularmente eficaz em tarefas como previsão de séries temporais, tradução automática, e processamento de linguagem natural (Graves, 2012).

Além disso, o modelo LSTM possui hiperparâmetros que são usados para definir as características do modelo. Eles desempenham um papel fundamental no desempenho da rede, pois controlam aspectos como a estrutura do modelo, a dinâmica do aprendizado e a regularização. Diferentemente dos pesos e vieses, que são ajustados pelo modelo durante o treinamento, os hiperparâmetros precisam ser definidos manualmente ou otimizados por meio de testes.

Em redes LSTM (Long Short-Term Memory), os hiperparâmetros são usados em diferentes componentes da arquitetura e do processo de treinamento:

1. **Arquitetura do Modelo:** Envolve parâmetros que definem a estrutura do modelo, como o número de camadas, o número de neurônios em cada camada e a função de ativação. Esses elementos determinam a capacidade da rede de modelar padrões complexos e não lineares.
2. **Processo de Otimização:** Inclui hiperparâmetros que influenciam como o modelo ajusta seus pesos durante o treinamento, como a taxa de aprendizado, o otimizador e o número de épocas. Esses parâmetros afetam tanto a velocidade quanto a estabilidade do processo de aprendizagem.
3. **Regularização:** Métodos como dropout, que se refere à uma técnica de regularização que "desativa" aleatoriamente uma fração de neurônios durante o treinamento, são usados para evitar *overfitting* e melhorar a capacidade do modelo de generalizar para novos dados.

4. **Processamento de Dados:** Parâmetros como o tamanho do lote (*batch size*) controlam como os dados são processados durante o treinamento, impactando o desempenho computacional e a precisão do modelo.

No presente estudo, foram analisados os seguintes hiperparâmetros do modelo LSTM, buscando encontrar uma configuração ideal para previsões mais precisas:

- **Neurônios por camada:** Os "neurônios" em cada camada da LSTM correspondem à quantidade de unidades LSTM presentes naquela camada. Cada unidade é capaz de "memorizar" informações temporais de forma seletiva.
- **Função de ativação:** Determina como os dados são transformados dentro das camadas ocultas, influenciando o comportamento não linear do modelo, permitindo que o modelo capture padrões mais complexos.
- **Otimizador:** O otimizador é um algoritmo que ajusta os pesos da rede neural durante o treinamento, com o objetivo de minimizar a função de perda (o erro entre as previsões do modelo e os valores reais).
- **Taxa de aprendizado:** Controla a magnitude dos ajustes nos pesos durante o treinamento.
- **Número de camadas:** Configura a profundidade da rede, adicionando complexidade para modelar padrões temporais.
- **Batch size:** O batch size é o número de amostras que o modelo processa antes de atualizar os pesos durante o treinamento. O treinamento não ocorre amostra por amostra, mas em pequenos grupos (*batches*), balanceando a eficiência computacional e a estabilidade.
- **Épocas:** As épocas (epochs) representam o número de vezes que o modelo passa por todo o conjunto de dados durante o treinamento. A cada época, o modelo ajusta os pesos do modelo com base nos erros das previsões.

2.6 ARIMA

O modelo ARIMA é um método estatístico amplamente reconhecido e utilizado para análise e previsão de séries temporais. Este modelo é capaz de capturar diversas estruturas temporais padronizadas em dados de séries temporais, oferecendo uma abordagem simplificada para realizar previsões eficazes dessas séries.

Conforme destacado por Hyndman e Athanasopoulos (2018), o modelo ARIMA é aplicado em situações em que os dados apresentam evidências de não estacionariedade, incorporando o conceito de integração para remover essa característica. Assim, o ARIMA

surge como uma generalização do modelo ARMA (*AutoRegressive Moving Average*), sendo ambos ajustados aos dados da série temporal com o objetivo de prever valores futuros.

O componente AR (*AutoRegressive*) do modelo descreve a relação de dependência entre uma observação e várias observações passadas. O modelo AR(p) pode ser descrito como: $y_t = \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \dots + \phi_p \cdot y_{t-p} + \varepsilon_t$. Onde c é o intercepto, ϕ é o coeficiente calculado para a defasagem, y_{t-n} a defasagem de y para n períodos anteriores, e ε_t é o erro aleatório do modelo, e p o número de defasagens utilizadas.

Já o MA (*Moving Average*) trata da dependência entre uma observação e o erro residual de um modelo de média móvel aplicado a observações anteriores. E pode ser descrito pela seguinte equação: $y_t = \theta_1 \cdot \varepsilon_{t-1} + \theta_2 \cdot \varepsilon_{t-2} + \dots + \theta_q \cdot \varepsilon_{t-q}$. Onde ε_t é o erro aleatório no tempo t , $\theta_1, \theta_2, \theta_q$ os coeficientes de média móvel atribuídos ao erro passado, e q o número de defasagens incluídas no modelo.

O I (*Integrated*) refere-se à utilização da diferenciação das observações brutas para tornar a série estacionária. Se d for 1, usaremos somente a primeira diferença:

$y_t^{(1)} = y_t - y_{t-1}$. Se d for 2, usaremos a segunda diferença: $y_t^{(2)} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$.

O modelo ARIMA é uma combinação desses 3 modelos AR, I e MA, e pode ser obtido a partir da seguinte estrutura: $z_t = (\text{Componente AR}) + (\text{Componente MA})$, sendo $z_t = \Delta y_t^d$. Por exemplo, um modelo ARIMA (1,1,1) pode ser representado da seguinte forma: $y_t - y_{t-1} = \phi_1 \cdot (y_{t-1} - y_{t-2}) + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1}$

Para que o modelo ARIMA forneça resultados satisfatórios, é essencial que os parâmetros de cada um de seus componentes sejam definidos antecipadamente (Box *et al.*, 1976). Esses parâmetros dependem do comportamento específico da série temporal em análise e devem ser testados a fim de encontrar a melhor combinação de parâmetros possível.

2.7 Validação dos modelos

Para validar os modelos, será utilizada a técnica de validação cruzada (*cross-validation*), amplamente empregada em *machine learning*. Os dados serão divididos em k grupos (*folds*), respeitando sua dependência temporal. Nesse formato, o grupo 1 sempre conterá os dados mais antigos, enquanto o grupo k representará os dados mais recentes.

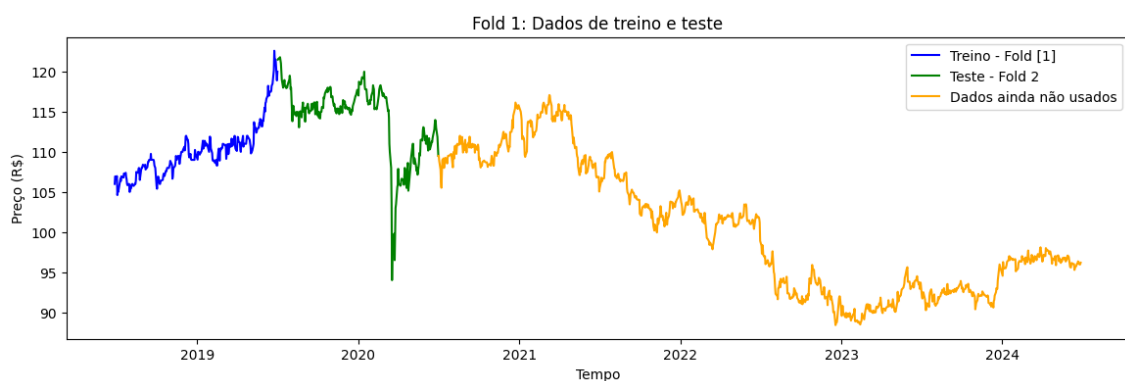
A abordagem específica utilizada será o método **k-fold cross-validation**, como descrito por Hastie, Tibshirani e Friedman (2009), garantindo que a avaliação dos modelos considere de forma adequada as características sequenciais da série temporal.

Nesse método, os k subconjuntos serão submetidos a treinamentos e testes de forma iterativa. Supondo um modelo subdividido em 6 folds, temos a seguinte estrutura de treinamento e validação do modelo:

- Primeira etapa: o modelo é treinado usando os dados do fold 1 (mais antigos) e testado no fold 2.
- Segunda etapa: o treinamento utiliza os dados dos folds 1 e 2, e o teste ocorre no fold 3.
- Terceira etapa: o treinamento abrange os folds 1, 2 e 3, enquanto o teste é feito no fold 4.
- Esse processo continua até a última etapa, onde o modelo é treinado com os folds de 1 a 5 e avaliado no fold 6 (mais recente).

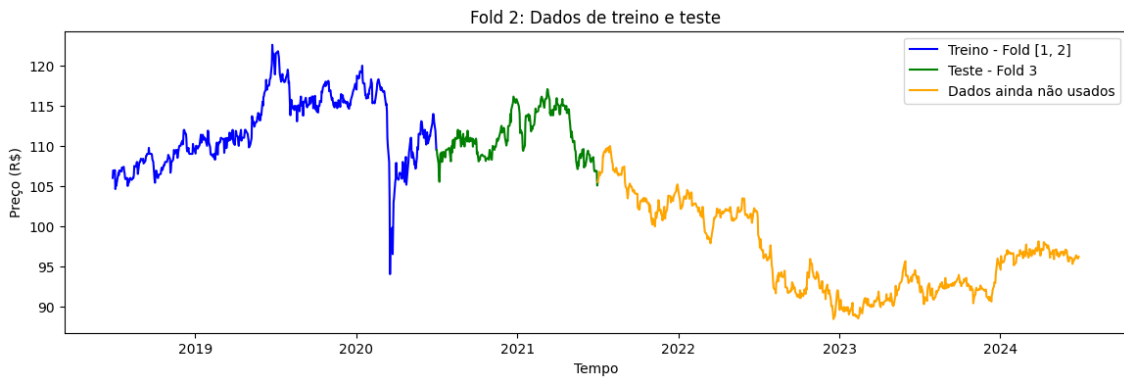
Nos gráficos 2,3,4,5,6, é apresentado um exemplo da divisão dos dados da ação KNIP11 em *folds*, de forma que em verde estão os dados usados para treino, em amarelo os dados que foram usados para teste, e em azul os dados que ainda não foram utilizados.

Figura 4 - 1ª Etapa do processo de k-fold cross-validation



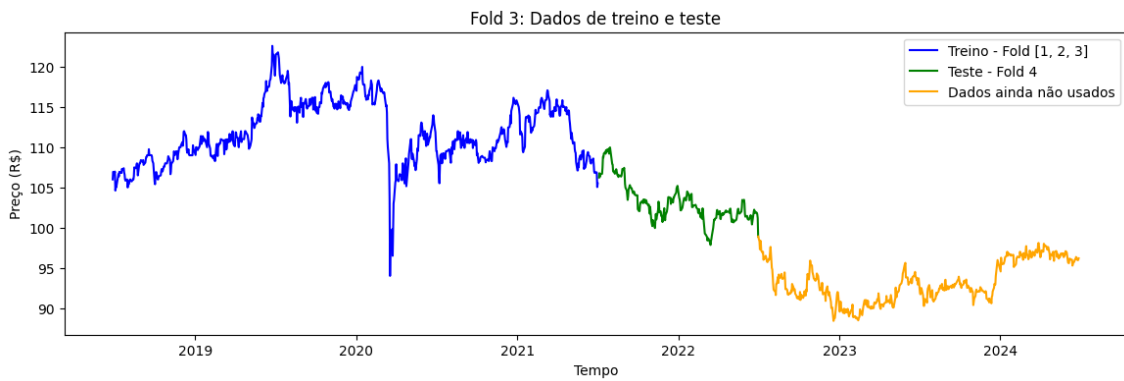
Fonte: Elaboração própria, 2024

Figura 5 - 2ª Etapa do processo de k-fold cross-validation



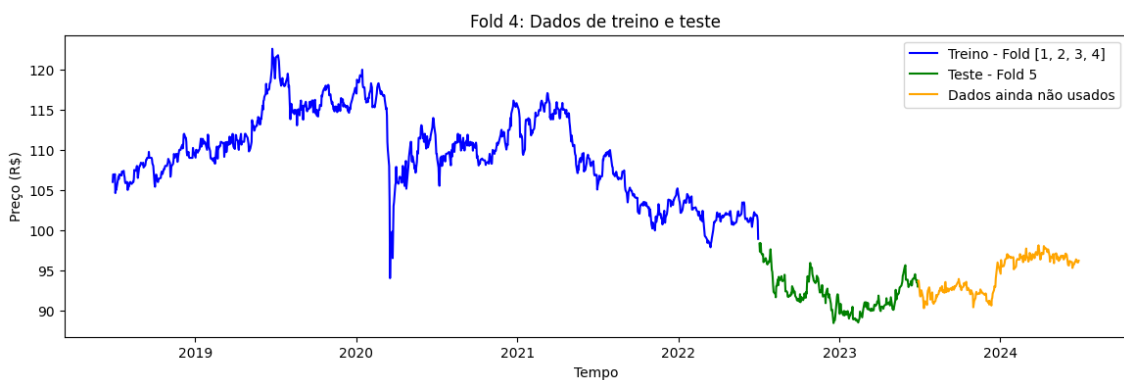
Fonte: Elaboração própria, 2024

Figura 6 - 3ª Etapa do processo de k-fold cross-validation



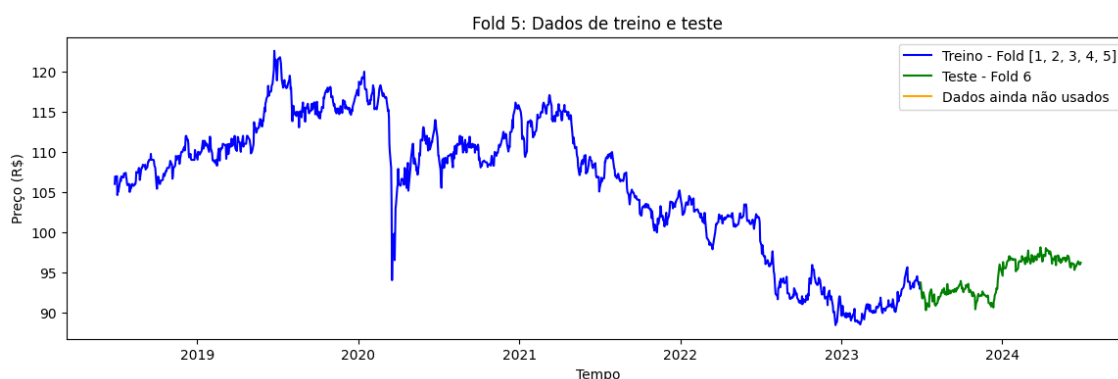
Fonte: Elaboração própria, 2024

Figura 7 - 4ª Etapa do processo de k-fold cross-validation



Fonte: Elaboração própria, 2024

Figura 8 - 5ª Etapa do processo de k-fold cross-validation



Fonte: Elaboração própria, 2024

Os modelos serão avaliados através da medida de erro MSE (*Mean Squared Error*). De acordo com Hyndman e Athanasopoulos (2018), a simplicidade e a robustez do MSE o caracterizam uma métrica-padrão em diversos campos da estatística e modelagem preditiva, permitindo uma avaliação consistente e comparável entre diferentes modelos. O MSE é medido através da fórmula: $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$. Onde n é o número total de observações, y_i representa o valor observado no instante i , \hat{y}_i é o valor previsto pelo modelo para o mesmo instante.

O MSE será medido para cada *fold* e no fim do processo de previsão, será tirado uma média do erro de cada *fold* e essa será a medida final de avaliação dos modelos.

2.8 Discussão Teórica sobre previsão de preços no Mercado Financeiro

Estudos recentes apontam as limitações de modelos clássicos na previsão de preços no mercado financeiro. Ling *et al.* (2020) destacaram que modelos clássicos de regressão linear não são eficazes na previsão de retornos de REITs (fundos imobiliários americanos), devido à complexidade intrínseca desses ativos. De forma semelhante, Zhang *et al.* (2023) demonstraram que modelos de *machine learning*, como o GMDH (Group Method of Data Handling), superam métodos tradicionais, como o GARCH, na previsão de fundos imobiliários.

O modelo ARIMA tem se mostrado promissor para previsões de curto prazo. Adebisi (2014) apresentou um processo detalhado de construção de modelos ARIMA para previsão de preços de ações, destacando a eficácia desses modelos para capturar padrões de curto prazo no mercado financeiro. Da mesma forma, Machado e Correa

(2022) utilizaram modelos ARIMA e LSTM para prever preços de ações, evidenciando a possibilidade de alcançar maior precisão na previsão de ativos financeiros.

Modelos de Redes Neurais Artificiais (RNA) têm sido amplamente explorados no mercado financeiro. Lertwachara (2007) comparou o desempenho de uma carteira escolhida aleatoriamente com uma construída a partir de uma rede neural ANN, mostrando um aumento de 26,05% nos resultados da carteira otimizada.

O uso de técnicas de *machine learning* no mercado financeiro está bem estabelecido na literatura. Silveira (2021) explorou a aplicação de Random Forest e LSTM no mercado de ações brasileiro, ressaltando o potencial dessas abordagens. Nelson (2017), por sua vez, comparou o desempenho do modelo LSTM com MLP (Multi-Layer Perceptron) e Random Forest, concluindo que o LSTM superou os outros métodos na previsão de preços no mercado brasileiro.

Adicionalmente, outras abordagens como Algoritmos Genéticos (Allen e Karjalainen, 1999), Máquinas de Vetor de Suporte (Jae Kim, 2003) e Prophet (Kadam et al., 2022) também são citadas como alternativas robustas para captura de padrões em séries temporais financeiras. Kamijo e Tanigawa (1990) propuseram o uso da própria série histórica de preços, enquanto Wang e Leu (1996) defenderam o uso de indicadores técnicos como variáveis independentes, como inflação e taxa de juros, para prever tendências.

Akita et al. (2016) destacaram o uso de LSTM para integrar dados históricos de preços com análise de sentimentos extraída de notícias financeiras. O modelo considerava o impacto do tom negativo das notícias na previsão de preços, adicionando uma variável representando a negatividade ao modelo. Chen et al. (2015) também utilizaram LSTM em combinação com índices de mercado, como o IBOVESPA, para prever tendências de preços no dia seguinte.

No contexto financeiro, as LSTMs têm se mostrado especialmente úteis na previsão de preços de ativos e na modelagem de séries temporais financeiras, onde a capacidade de capturar e processar padrões complexos e não lineares é crucial. Estudos como o de Fischer e Krauss (2018) demonstraram que as LSTMs podem superar modelos tradicionais, como ARIMA, ao lidar com a natureza volátil e dinâmica dos mercados financeiros.

3 METODOLOGIA

Esta Seção detalha a metodologia adotada para a coleta de dados, pré-processamento, treinamento e validação dos modelos e quais análises serão feitas. Serão explicados os critérios de seleção dos fundos imobiliários, as etapas do pré-processamento aplicado e como os modelos serão treinados e avaliados.

Inicialmente, serão apresentados os critérios de seleção da base de dados utilizada, bem como o tratamento aplicado aos dados para garantir sua adequação aos modelos preditivos. Em seguida, será descrita a construção dos modelos LSTM e ARIMA, incluindo a definição dos hiperparâmetros e as técnicas empregadas para otimização e validação dos resultados. Além disso, serão explicadas as métricas utilizadas para avaliar o desempenho de cada modelo, com o objetivo de comparar sua efetividade na previsão dos preços de fechamento dos fundos imobiliários analisados.

3.1 Base de Dados

Serão utilizados dados secundários fornecidos pelo *Yahoo Finance*, que captura os dados diariamente de todas as ações listadas em bolsa, registrando o preço mínimo, máximo, médio e volume. A priori, foi utilizada a biblioteca ‘tradingcomdados’, que captura dados da B3 (bolsa de valores brasileira) e fornece para os usuários informações importantes, como a composição do IFIX. Dessa forma, foi elaborada uma tabela com o nome dos fundos, e sua participação no IFIX.

Na tabela 1 é apresentado os 5 primeiros fundos, onde ‘Cód_Fundo’ é o código do fundo que é negociado na bolsa e ‘Participação_IFIX’ sua participação no índice IFIX em porcentagem.

Tabela 1 - Informações sobre os fundos que compõem o IFIX.

Cód_Fundo	Nome_Fundo	Participação_IFIX
CACR11	CARTESIA FII	0,382%
BPF11	FII ABSOLUTO	0,222%
AFHI11	FII AFHI CRI	0,360%
AJFI11	FII AJ MALLS	0,189%
ALZR11	FII ALIANZA	1,018%

Fonte: Elaboração própria, 2024.

Com todos os códigos dos fundos que compõem o IFIX, foi realizada uma vinculação entre o código do ativo e o CNPJ do fundo de investimento. Esse

procedimento foi realizado através do site <https://www.fundsexplorer.com.br>, que fornece informações gerais sobre os FII's, onde cada fundo foi consultado e teve seu CNPJ vinculado. Através do CNPJ de cada fundo, foi consultado na CVM (Comissão de Valores Mobiliários) seu histórico, e assim, foi obtido a data do início do fundo, seu patrimônio líquido e o setor de atuação.

Tabela 2 - Dados completos dos fundos que compõem o IFIX.

Cód_Fundo	ID_CNPJ_Fundo	Data_Início	Setor	Valor_Patrimônio
CACR11	32.065.364/0001-46	17/10/2019	Papéis	R\$ 381.793.035,89
BPFF11	17.324.357/0001-28	09/04/2013	Fundo de Fundos	R\$ 348.481.390,33
AFHI11	36.642.293/0001-58	03/03/2021	Papéis	R\$ 382.833.867,39
AJFI11	51.472.985/0001-99	01/11/2023	Shoppings	R\$ 394.707.426,45
ALZR11	28/737.771/0001-85	27/12/2017	Misto	R\$ 1.044.422.993,15

Fonte: Elaboração própria, 2024.

Na tabela 2, há a base de dados de 5 fundos que compõem o IFIX, juntamente com seu CNPJ, sua data de início de atividades, seu setor de atuação e seu patrimônio líquido. Essa será a base de dados que será usada para selecionar os fundos que serão utilizados para a previsão de preços.

Os fundos foram subdivididos nos 5 tipos de FII determinados pela ANBIMA da seguinte forma:

- **Fundos de papel:** Incluem os fundos do setor de atuação "Papéis".
- **Fundos de tijolo:** Compreendem os fundos dos setores "Hotéis", "Imóveis Comerciais – Outros", "Imóveis Industriais e Logísticos", "Lajes Corporativas", "Shoppings" e "Varejo".
- **Fundos de desenvolvimento:** Agrupam os fundos do setor "Fundo de Desenvolvimento".
- **Fundos híbridos:** Representam os fundos do setor "Misto".
- **Fundos de fundos:** Incluem os fundos do setor "Fundo de Fundos".

A partir dessa subdivisão, foram selecionados os maiores fundos ranqueados pelo patrimônio líquido para cada tipo de fundo definido pela ANBIMA.

Os selecionados foram:

Tabela 3 - Informações sobre os fundos escolhidos

Cód_Fundo	ID_CNPJ_Fundo	Data_Início	Setor	Valor_Patrimônio
KNIP11	24.960.430/0001-13	16/09/2016	Papéis	R\$ 7.619.242.765,39
HGLG11	11.728.688/0001-47	06/05/2010	Imóveis Industriais e Logísticos	R\$ 5.306.763.276,83
MFII11	16.915.968/0001-88	06/03/2010	Fundos de Desenvolvimento	R\$ 570.350.300,55
KNRI11	16.915.968/0001-88	22/10/2010	Misto	R\$ 4.221.162.551,49
HFOF11	16.915.968/0001-88	27/02/2018	Fundo de Fundos	R\$ 1.821.118.103,96

Fonte: Elaboração própria, 2024.

Dessa forma, os fundos que serão utilizados para a avaliação dos modelos de previsão são: KNIP11 (Fundo de Papel), HGLG11 (Fundo de Tijolo), MFII11 (Fundo de desenvolvimento), KNRI11 (Fundo Híbrido), HFOF11 (Fundo de fundos).

O histórico de preços de cada fundo foi obtido através do Yahoo Finanças (<https://br.financas.yahoo.com>). A partir do histórico de preços, foi definido uma data de corte para a utilização desses dados, a data de início será 28/06/2018 e a data de fim será 28/06/2024, compondo um período de 6 anos.

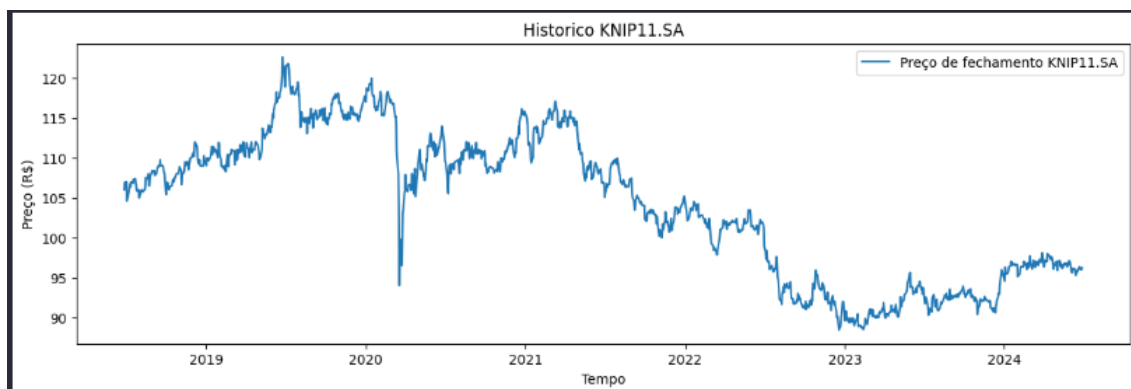
Tabela 4 - Exemplo de histórico de preços do fundo KNIP11.

Data	Abertura	Maior	Menor	Fechamento	Volume
28/06/2018	105,98	107,00	105,05	106,00	12.928
29/06/2018	106,00	106,95	106,00	106,94	7.875
02/07/2018	108,00	108,00	106,00	106,75	3.625
03/07/2018	106,75	107,69	106,00	106,97	12.319
04/07/2018	107,00	107,00	104,61	104,61	10.367

Fonte: Elaboração própria, 2024

A Tabela 4 apresenta os cinco primeiros registros históricos do fundo imobiliário KNIP11, contendo as seguintes informações: a coluna Abertura, que indica o preço de abertura do ativo naquele dia; a coluna Maior, que registra o maior preço alcançado no dia; a coluna Menor, com o menor preço registrado; a coluna Fechamento, representando o preço de fechamento do dia; e, por fim, a coluna Volume, que indica a quantidade total de negociações realizadas ao longo do dia.

Figura 9 - Série completa de preços de fechamento do KNIP11



Fonte: Elaboração própria, 2024

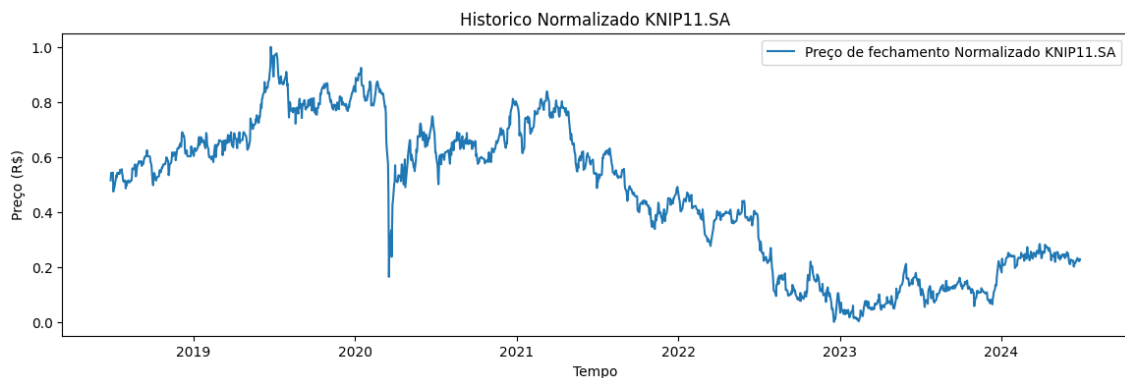
No gráfico 7 é apresentada uma representação gráfica do histórico dos preços de fechamento do KNIP11, objeto de estudo deste trabalho.

3.2 Tratamento dos dados

A partir da base de dados obtida do Yahoo Finance, a coluna Fechamento foi armazenada, já que será o objeto de previsão. Conforme realizado no trabalho de (SANTOS, 2019), os dados foram normalizados utilizando a seguinte fórmula: $x_t = \frac{x_t - x_{min}}{(x_{max} - x_{min})}$.

Essa transformação se faz necessária a fim de facilitar o processamento dos dados pelos modelos, já que irá ajudar na convergência do otimizador. Com essa normalização, os dados passam a variar entre 0 e 1, de forma que os cálculos realizados pelos modelos para atingir o melhor modelo são realizados de forma mais eficiente, (JAMES *et al.*, 2018).

Figura 10 - Série completa de preços de fechamento normalizados do KNIP11



Na figura 10, é demonstrada a normalização do preço de fechamento do FII KNIP11, de forma que desse modo, os preços de fechamento agora variam de 0 a 1.

3.3 Construção dos modelos

3.3.1 Modelo LSTM

O modelo LSTM foi elaborado para realizar previsões dos preços de fechamento dos fundos utilizando uma abordagem de janela deslizante com três configurações distintas: 10, 30 e 60 variáveis passadas.

Para treinar e avaliar o modelo LSTM através do python, foi utilizada a biblioteca *tensorflow*, que possui a capacidade para definição dos hiperparâmetros do modelo. Para definir o melhor modelo de LSTM para cada série temporal, foram testadas as seguintes variações de hiperparâmetros no modelo:

Tabela 5 - Hiperparâmetros LSTM Testados

Hiperparâmetro	Valores Testados
Neurônios	50,100
Função de Ativação	ReLU, Tanh
Otimizador	Adam, SGD
Taxa de Aprendizado	0.001, 0.01
Taxa de Dropout	0,0, 0,2
Número de Camadas	1, 2
Batch Size	16, 32
Épocas	50, 100

Fonte: Elaboração própria, 2024.

Dessa forma, 256 modelos distintos foram avaliados para cada combinação de fundo e janela deslizante, totalizando uma análise abrangente e detalhada.

Os modelos foram treinados e avaliados utilizando a técnica de validação cruzada, respeitando a dependência temporal da série. O desempenho de cada configuração foi mensurado através do erro médio quadrático (**MSE**), permitindo identificar a combinação de hiperparâmetros e janela deslizante que apresentou o menor erro preditivo.

O modelo LSTM com o menor MSE para cada fundo foi então selecionado para comparação direta com o modelo ARIMA.

3.3.2 Modelo ARIMA

O modelo ARIMA foi desenvolvido com premissas semelhantes às adotadas para o modelo LSTM, visando prever os preços de fechamento. Para isso, foram utilizadas janelas deslizantes contendo 10, 30 e 60 observações passadas como base para as previsões.

Tabela 6 Hiperparâmetros ARIMA Testados

Hiperparâmetro	Valores Testados
p (autoregressão)	0, 1, 2, 3
d (diferenciação)	0, 1, 2, 3
q (média móvel)	0, 1, 2, 3

Fonte: Elaboração própria, 2024.

Para ajustar o modelo, foram testadas as combinações possíveis dos hiperparâmetros p, d e q, variando de 0 a 3, totalizando 64 modelos distintos para cada quantidade de janelas deslizantes. Cada combinação de hiperparâmetros foi avaliada em termos de desempenho preditivo, utilizando validação cruzada, respeitando a dependência temporal dos dados, de forma equivalente a usada para avaliar o modelo LSTM.

Ao final do processo, as configurações de hiperparâmetros e janela deslizante que apresentaram o menor erro médio quadrático (**MSE**) foram selecionadas como as melhores combinações para o modelo ARIMA, permitindo a comparação direta com o modelo LSTM. Essa abordagem assegura que tanto o LSTM quanto o ARIMA sejam avaliados em suas configurações mais eficientes, proporcionando uma análise justa e fundamentada de suas capacidades preditivas.

4 AVALIAÇÃO E DISCUSSÃO DOS RESULTADOS

Nesta seção, serão apresentados e discutidos os resultados obtidos pelos modelos LSTM e ARIMA para cada ativo analisado. A análise será realizada considerando, inicialmente, o desempenho de ambos os modelos, baseado na métrica de avaliação apresentada, seguido da identificação do melhor método para previsão em cada ativo, com base nas métricas avaliadas. Além disso, serão destacados os melhores parâmetros selecionados para cada modelo, bem como a janela deslizante que apresentou maior eficácia na previsão.

4.1 Resultados obtidos com o fundo HFOF11

Tabela 7 - Erros medidos pelos modelos LSTM e ARIMA para o fundo HFOF11.SA

Janela Deslizante	LSTM	ARIMA
10	0,8897	0,6821
30	1,2514	0,6821
60	0,6965	0,6821

Fonte: Elaboração Própria, 2024

A análise da Tabela 7 revela a comparação entre os melhores modelos LSTM e ARIMA, com os erros de previsão (MSE) avaliados para o fundo HFOF11.SA em três configurações de janelas deslizantes: 10, 30 e 60 observações. Os resultados apresentados demonstram que o modelo ARIMA foi o mais eficiente para todas as janelas deslizantes analisadas, com erro constante de 0,6821. O modelo ARIMA foi, em média, 23,06% mais eficiente que o modelo LSTM.

Tabela 8 - Melhores modelos LSTM – HFOF11

Janela Deslizante \ Hiperparâmetros	10	30	60
Neurônios	100	50	100
Função de Ativação	Tanh	Tanh	Tanh
Otimizador	Adam	Adam	Adam
Taxa de Aprendizado	0.01	0.01	0.01
Taxa de Dropout	0.0	0.0	0.0
Número de Camadas	1	1	1
Batch Size	16	16	16
Épocas	100	50	100

Fonte: Elaboração Própria, 2024

Tabela 9 - Melhores modelos ARIMA – HFOF11

Janela Deslizante \ Hiperparâmetros	10	30	60
p	0	0	0
d	1	1	1
q	0	0	0

Fonte: Elaboração Própria, 2024

As Tabelas 8 e 9 apresentam os melhores modelos identificados para os métodos LSTM e ARIMA no fundo HFOF11, detalhando os hiperparâmetros ajustados para cada janela deslizante analisada. Esses resultados justificam os erros obtidos anteriormente, evidenciando as configurações que impactaram no desempenho dos modelos.

O modelo ARIMA (0,1,0), que representa um Random Walk, foi robusto e apresentou erros constantes e baixos independentemente do tamanho da janela deslizante. Esse comportamento reforça a aleatoriedade encontrada na definição dos preços desse fundo, já que um modelo de passeio aleatório conseguiu prever melhor em comparação a um modelo de *deep learning* otimizado para aquela série temporal.

4.2 Resultados obtidos com o fundo HGLG11

Tabela 10 - Erros medidos pelo melhor modelo LSTM e ARIMA para o fundo HGLG11.SA

Janela Deslizante	LSTM	ARIMA
10	4,2385	3,7481
30	5,2289	3,7481
60	4,0243	3,7481

Fonte: Elaboração Própria, 2024

Os resultados obtidos para os melhores modelos LSTM e ARIMA, explicitados na tabela 7, demonstram que o modelo ARIMA foi superior para todas as janelas deslizantes, como também ocorreu para o HFOF11, com um MSE constante de 3,7481. O modelo ARIMA foi, em média, 15,58% mais eficiente que o modelo LSTM.

Tabela 11 - Melhores modelos LSTM – HGLG11

Janela Deslizante	10	30	60
Hiperparâmetros			
Neurônios	100	50	50
Função de Ativação	ReLU	Tanh	ReLU
Otimizador	Adam	Adam	Adam
Taxa de Aprendizado	0.01	0.01	0.01
Taxa de Dropout	0.0	0.0	0.0
Número de Camadas	1	1	1
Batch Size	16	16	16
Épocas	100	50	100

Fonte: Elaboração Própria, 2024

Tabela 12 - Melhores modelos ARIMA - HGLG11

Janela Deslizante \ Hiperparâmetros	10	30	60
p	0	0	0
d	1	1	1
q	0	0	0

Fonte: Elaboração Própria, 2024

Semelhante ao que foi visto no HFOF11, o melhor modelo ARIMA encontrado para previsão de preços do HGLG11, foi o ARIMA (0,1,0) para todas as variações de janelas deslizantes. Isso demonstra que os preços desse fundo também um passeio aleatório, onde os preços futuros são definidos apenas pelo preço atual adicionado de um ruído.

4.3 Resultados obtidos com o fundo KNIP11

Tabela 13 - Erros medidos pelo melhor modelo LSTM e ARIMA para o fundo
KNIP11.SA

Janela Deslizante	LSTM	ARIMA
10	0,7449	0,5745
30	0,7208	0,5745
60	1,1846	0,5745

Fonte: Elaboração Própria, 2024

Os resultados obtidos para os melhores modelos LSTM e ARIMA, explicitados na tabela 13, demonstram que o modelo ARIMA foi superior para todas as janelas deslizantes, como também ocorreu para o HFOF11, com um MSE constante de 0,5745. O modelo ARIMA foi, em média, 31,56% mais eficiente que o modelo LSTM.

Tabela 14 - Melhores modelos LSTM - KNIP11

Janela Deslizante \ Hiperparâmetros	10	30	60
Neurônios	50	50	100
Função de Ativação	Tanh	Tanh	Tanh
Otimizador	Adam	Adam	Adam
Taxa de Aprendizado	0.01	0.01	0.01
Taxa de Dropout	0.0	0.0	0.0
Número de Camadas	1	1	1
Batch Size	16	16	16
Épocas	50	50	50

Fonte: Elaboração Própria, 2024

Tabela 15 - Melhores modelos ARIMA - KNIP11

Janela Deslizante \ Hiperparâmetros	10	30	60
p	0	0	0
d	1	1	1
q	0	0	0

Fonte: Elaboração Própria, 2024

Como pode ser observado na tabela 15, o melhor modelo ARIMA encontrado para previsão de preços do KNIP11, foi o ARIMA (0,1,0) para todas as variações de janelas deslizantes. Isso demonstra que os preços desse fundo também um passeio aleatório, onde os preços futuros são definidos apenas pelo preço atual mais um ruído.

4.4 Resultados obtidos com o fundo KNRI11

Tabela 16 - Erros medidos pelo melhor modelo LSTM e ARIMA para o fundo KNRI11.SA

Janela Flutuante	LSTM	ARIMA
10	3,0611	2,6404
30	3,8405	2,6404
60	3,5355	2,6404

Fonte: Elaboração Própria, 2024

Os resultados obtidos para os melhores modelos LSTM e ARIMA, explicitados na tabela 16, demonstram que o modelo ARIMA foi superior para todas as janelas deslizantes, como também ocorreu para o HFOF11, com um MSE constante de 2,6404. O modelo ARIMA foi, em média, 23,44% mais eficiente que o modelo LSTM.

Tabela 17 - Melhores modelos LSTM - KNRI11

Janela Deslizante	10	30	60
Hiperparâmetros			
Neurônios	100	50	100
Função de Ativação	Tanh	Tanh	Tanh
Otimizador	Adam	Adam	Adam
Taxa de Aprendizado	0.01	0.001	0.001
Taxa de Dropout	0.0	0.0	0.2
Número de Camadas	1	2	1
Batch Size	16	16	16
Épocas	100	100	100

Fonte: Elaboração Própria, 2024

Tabela 18 - Melhores modelos ARIMA - KNRI11

Janela Deslizante \ Hiperparâmetros	10	30	60
p	0	0	0
d	1	1	1
q	0	0	0

Fonte: Elaboração Própria, 2024

Como pode ser observado na tabela 18, o melhor modelo ARIMA encontrado para previsão de preços do KNRI11, foi o ARIMA (0,1,0) para todas as variações de janelas deslizantes. Isso demonstra que os preços desse fundo também um passeio aleatório, onde os preços futuros são definidos apenas pelo preço atual mais um ruído.

4.5 Resultados obtidos com o fundo MFII11

Tabela 19 - Erros medidos Erros medidos pelo melhor modelo LSTM e ARIMA para o fundo MFII11.SA

Janela Flutuante	LSTM	ARIMA
10	1,1795	1,1441
30	2,0990	1,1441
60	3,4966	1,1441

Fonte: Elaboração Própria, 2024

Os resultados obtidos para os melhores modelos LSTM e ARIMA, explicitados na tabela 19, demonstram que o modelo ARIMA foi superior para todas as janelas deslizantes, como também ocorreu para o HFOF11, com um MSE constante de 1,1441. O modelo ARIMA foi, em média, 38,59% mais eficiente que o modelo LSTM.

Tabela 20 - Melhores modelos LSTM - MFII11

Janela Deslizante \ Hiperparâmetros	10	30	60
Neurônios	100	50	100
Função de Ativação	Tanh	Tanh	Tanh
Otimizador	Adam	Adam	Adam
Taxa de Aprendizado	0.01	0.01	0.01
Taxa de Dropout	0.0	0.2	0.2
Número de Camadas	1	2	2
Batch Size	16	32	32
Épocas	100	100	100

Fonte: Elaboração Própria, 2024

Tabela 21 - Melhores modelos ARIMA - MFII11

Janela Deslizante \ Hiperparâmetros	10	30	60
p	0	0	0
d	1	1	1
q	0	0	0

Fonte: Elaboração Própria, 2024

Como pode ser observado na tabela 21, o melhor modelo ARIMA encontrado para previsão de preços do MFII11, foi o ARIMA (0,1,0) para todas as variações de janelas deslizantes. Isso demonstra que os preços desse fundo também um passeio aleatório, onde os preços futuros são definidos apenas pelo preço atual mais um ruído.

5 CONCLUSÃO

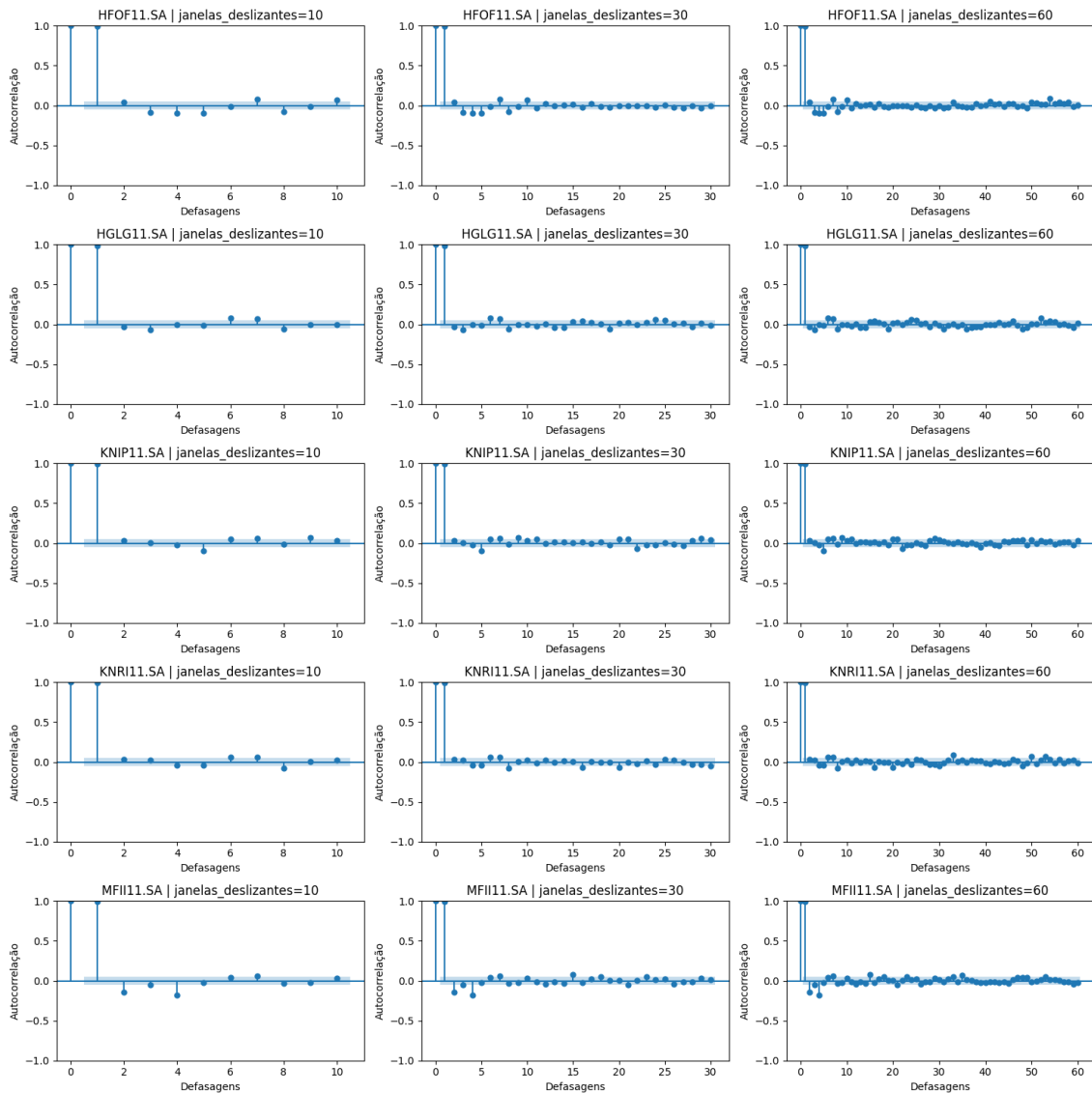
No contexto do mercado financeiro, este trabalho abordou a previsão de preços de fechamento de fundos imobiliários utilizando os modelos LSTM e ARIMA para diferentes variações de janelas deslizantes. Devido à relevância do mercado financeiro, é fundamental compreender a formação e previsão dos preços desses ativos, uma vez que desempenham um papel importante na diversificação de portfólios e na gestão de riscos dos investidores.

O mercado de fundos imobiliários apresenta características de volatilidade e alta sensibilidade às condições macroeconômicas e setoriais. Para mitigar tais riscos e facilitar o processo de tomada de decisão, este estudo buscou construir modelos preditivos capazes de gerar algum nível de previsibilidade para os preços de fechamento desses fundos.

Por meio de técnicas de *machine learning*, como o LSTM, e do tradicional modelo ARIMA, foram realizadas projeções para diferentes janelas deslizantes (10, 30 e 60 períodos). Os resultados obtidos indicaram que o modelo ARIMA (0,1,0) apresentou menor erro e maior robustez na previsão dos preços em comparação ao modelo LSTM, em todas as janelas analisadas. O ARIMA (0,1,0), que representa um Random Walk, reflete a ideia de que os preços seguem um comportamento aleatório ao longo do tempo, sem uma tendência ou padrão claro que possa ser capturado por modelos mais complexos.

Embora o LSTM seja conhecido por sua capacidade de capturar dependências de longo prazo em séries temporais, seu desempenho pode ter sido prejudicado pelo caráter essencialmente aleatório e pela alta volatilidade dos preços dos fundos imobiliários analisados. Quando se plota os gráficos de autocorrelação das séries temporais analisadas, é possível verificar que só há correlação entre o preço do dia exatamente anterior.

Figura 11 - Função de autocorrelação das séries temporais analisadas.



Elaboração própria, 2024

Esse comportamento poderia indicar que há pelo menos uma autocorrelação válida, possibilitando o uso do ARIMA (1,0,0). Mas ao verificar as estacionariedade das séries temporais, têm-se que elas não são estacionárias, sendo necessário então, diferenciar essa série até torná-la estacionária. Dessa forma, ao plotar novamente a função de autocorrelação, têm-se que os preços não tem nenhuma autocorrelação válida. Esse fato reforça a presença de um Random Walk nas séries analisadas.

Adicionalmente, ao comparar o desempenho médio entre os modelos, verificou-se que o ARIMA foi, em média, 26,56% mais eficiente do que o LSTM, considerando todas as janelas deslizantes e fundos imobiliários avaliados. Esses achados reforçam a eficácia de modelos mais simples e baseados em premissas estatísticas clássicas na

previsão de séries temporais com alta aleatoriedade, em oposição a métodos mais complexos de aprendizado de máquina.

Portanto, este estudo cumpriu o objetivo de comparar a capacidade preditiva dos modelos LSTM e ARIMA no mercado financeiro, trazendo insights relevantes sobre a aplicação de métodos computacionais no contexto dos fundos imobiliários.

6 TRABALHOS FUTUROS

Como trabalhos futuros, sugere-se a realização de estudos que explorem modelos capazes de capturar tendências nos preços dos fundos imobiliários, uma vez que a previsão direta de preços pode apresentar limitações, especialmente em séries temporais com comportamentos aleatórios, como observado neste estudo. Nesse sentido, o uso de modelos de *machine learning* para a identificação de padrões de tendência pode se mostrar mais promissor.

Trabalhos anteriores, como os de Chang *et al.* (2024), evidenciam que técnicas de aprendizado de máquina são mais eficazes na detecção de tendências e direções dos movimentos do mercado do que na previsão exata de valores de preços. Segundo os autores, a capacidade desses modelos de identificar padrões não lineares e sinais fracos em séries temporais complexas permite uma análise mais robusta e informativa para investidores e tomadores de decisão.

Além disso, seria interessante validar o desempenho dos modelos de tendência não apenas com métricas estatísticas, mas também por meio de resultados financeiros. Por exemplo, avaliar se os sinais gerados pelos modelos foram capazes de indicar pontos de compra e venda que resultassem em ganhos reais. A análise da rentabilidade acumulada ao longo de um período, considerando estratégias de negociação baseadas nos sinais preditivos, poderia fornecer uma avaliação prática e robusta da eficácia dos modelos, alinhando a pesquisa acadêmica com as demandas do mercado financeiro.

Outra abordagem possível envolve a decomposição da série temporal em seus componentes principais: tendência, sazonalidade e ruído aleatório. Essa técnica pode ajudar a entender melhor a estrutura dos dados e possibilitar previsões mais precisas ao modelar cada componente separadamente.

Estudos futuros também podem incluir a incorporação de variáveis exógenas, como indicadores macroeconômicos e índices setoriais, que podem auxiliar na identificação de tendências de longo prazo e no entendimento das dinâmicas que influenciam os preços dos fundos imobiliários.

Além disso, a utilização de modelos baseados em Transformers para previsão de séries temporais multivariadas pode ser uma abordagem promissora. Já que diferentemente das LSTMs, eles processam dados em paralelo e capturam relacionamentos entre todos os elementos em uma sequência simultaneamente.

7 REFERÊNCIAS

ABADI, Martín; AGARWAL, Ashish; BARHAM, Paul; BREVDO, Eugene; CHEN, Zhifeng; CITRO, Craig; CORRADO, Greg S.; DAVIS, Andy; DEAN, Jeffrey; DEVIN, Matthieu; GHEMAWAT, Sanjay; GOODFELLOW, Ian; HARP, Andrew; IRVING, Geoffrey; ISARD, Michael; JOSEFOWICZ, Rafal; JIA, Yangqing; KAISER, Lukasz; KUDLUR, Manjunath; LEVENBERG, Josh; MANÉ, Dan; SCHUSTER, Mike; MONGA, Rajat; MOORE, Sherry; MURRAY, Derek; OLAH, Chris; SHLENS, Jonathon; STEINER, Benoit; SUTSKEVER, Ilya; TALWAR, Kunal; TUCKER, Paul; VANHOUCHE, Vincent; VASUDEVAN, Vijay; VIÉGAS, Fernanda; VINYALS, Oriol; WARDEN, Pete; WATTENBERG, Martin; WICKE, Martin; YU, Yuan; ZHENG, Xiaoqiang. **TensorFlow: Large-scale machine learning on heterogeneous systems**. 2015. Disponível em: <https://www.tensorflow.org/>. Acesso em 30 jan. 2025.

ADEBIYI, Ayodele Ariyo; Adewumi, Aderemi; Ayo Charles. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. **Journal of Applied Mathematics**. v. 1, p 1-7, 2014. Disponível em: https://www.researchgate.net/publication/260821414_Comparison_of_ARIMA_and_Artificial_Neural_Networks_Models_for_Stock_Price_Prediction. Acesso em 15 ago. 2024.

AKITA, Ryo; YOSHIHARA, Akira; MATSUBARA, Takashi; Uehara, Kuniaki. Deep learning for stock prediction using numerical and textual information. IEEE/ACIS INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION SCIENCE, p. 1-6, 2016. Disponível em: <https://ieeexplore.ieee.org/document/7550882>. Acesso em 15 ago. 2024.

ALLEN, F.; KARJALAINEN, R. Using genetic algorithms to find technical trading rules. **Journal of Financial Economics**, v. 51 n. 2, p 245 – 271, 1991. Disponível em: https://econpapers.repec.org/article/eeeeifinec/v_3a51_3ay_3a1999_3ai_3a2_3ap_3a245-271.htm. Acesso em 15 ago. 2024.

ANBIMA. Classificação Fundos Imobiliários. 2015. Disponível em: https://www.anbima.com.br/pt_br/noticias/fundos-imobiliarios-ganham-classificacao-inedita-no-mercado.htm. Acesso em 15 ago. 2024.

AMORIM, L. T; CALLOU, Gustavo; PEREIRA, Saulo; SANTOS A.S. Using the LSTM Network for Stock Prediction in Several Economic Sectors of the Stock Exchange. **Revista Eletrônica de Gestão Organizacional**. v. 21, n.1 2023. Disponível em: <https://periodicos.ufpe.br/revistas/index.php/gestaoorg/article/view/257384>. Acesso em 31 ago. 2024.

BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. **IEEE Transactions on Neural Networks**. v. 5, p. 157-166, 1994. Disponível em: <https://ieeexplore.ieee.org/document/279181>. Acesso em 15 ago. 2024.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: Forecasting and Control**. 5. ed: Wiley. p. 189-210, 1976. Disponível em: [http://repo.darmajaya.ac.id/4781/1/Time%20Series%20Analysis_%20Forecasting%20and%20Control%20\(%20PDFDrive%20\).pdf](http://repo.darmajaya.ac.id/4781/1/Time%20Series%20Analysis_%20Forecasting%20and%20Control%20(%20PDFDrive%20).pdf). Acesso em: 11 ago. 2024.

BRASIL. Comissão de Valores Mobiliários. **Instrução CVM nº 472**, de 31 de outubro de 2008. Disponível em: <https://conteudo.cvm.gov.br/legislacao/instrucoes/inst472.html>. Acesso em: 15 ago. 2024.

CHANG, V.; XU, Q. A.; CHIDOZIE, A.; WANG, H. **Predicting Economic Trends and Stock Market Prices with Deep Learning and Advanced Machine Learning Techniques**. v. 13, n. 17, p. 3396, 2024. Disponível em: <https://www.mdpi.com/2079-9292/13/17/3396>. Acesso em: 15 dez. 2024

CHAVES, Marcos. *Códigos do TCC: Efetividade de Modelos de Previsão de Fundos Imobiliários*. 2024. Disponível em: https://github.com/sepectrok/codigos_tcc. Acesso em: 24 fev. 2025.

CHEN, K.; ZHOU, Y. e DAI, F. A LSTM-based method for stock returns prediction: A case study of china stock market. **IEEE International Conference** p. 2823–2824, 2015. Disponível em: <https://ieeexplore.ieee.org/document/7364089>. Acesso em 14 ago. 2024.

CHOLLET, François; et al. **Keras**. 2015. Disponível em: <https://github.com/keras-team/keras>. Acesso em 30 jan. 2025.

CICI G; CORGEL J; GIBSON S. Can fund managers select outperforming REITs? Examining fund holdings and trades. **Real Estate Econ**, 2011. Disponível em: https://www.researchgate.net/publication/227376872_an_Fund_Managers_Select_Outperforming_REITs_Examining_Fund_Holdings_and_Trades. Acesso em: 14 ago. 2024.

FAMA, Eugene F. Efficient capital markets: A review of theory and empirical work. **The Journal of Finance**, v. 25, n. 2, p 383-417, 1970. Disponível em: <https://www.jstor.org/stable/2325486>. Acesso em 15 ago. 2024.

FISCHER, T.; KRAUSS, C. Deep learning with long short-term memory networks for financial market predictions. **European Journal of Operational Research**. v. 270, n.2, p 654-669, 2018. Disponível em:

GEEKSFORGEEEKS. *Deep learning: introduction to long short-term memory*. Disponível em: <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>. Acesso em: 27 dez. 2024.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press. 2016. Disponível em: <http://www.deeplearningbook.org> Acesso em 15 ago. 2024.

GRAVES, A. Supervised Sequence Labelling with Recurrent Neural Networks, **Studies in Computational Intelligence**. Springer. v. 385, 2012. Disponível em: <https://link.springer.com/book/10.1007/978-3-642-24797-2>;. Acesso em 15 ago. 2024.

HAYKIN, Simon. **Neural Networks: A Comprehensive Foundation**. Prentice Hall. 1998. Disponível em: https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Neural%20Networks%20-%20A%20Comprehensive%20Foundation%20-%20Simon%20Haykin.pdf. Acesso em 15 ago. 2024.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735-1780, 1997. Disponível em:

<https://direct.mit.edu/neco/article-abstract/9/8/1735/6109/Long-Short-Term-Memory?redirectedFrom=fulltext>. Acesso em: 11 ago. 2024.

HYNDMAN, J. Rob; ATHANASOPOULOS, George. **Forecasting: Principles and Practice**. O Texts. 2ªed. 2018. Disponível em: <https://otexts.com/fpp2/index.html>. Acesso em 14 ago. 2024.

JAMES, G.; WITTEN, D.; HASTIE T.; TIBSHIRANI R. An Introduction to Statistical Learning: with Applications in R. 2018 Disponível em: <https://link.springer.com/book/10.1007/978-3-031-38747-0>. Acesso em 15 dez. 2024.

KADAM, Y., Kulkarni, S., Lonsane, S. & Khandagale, A.S. A Survey on Stock Market Price Prediction System using Machine Learning Techniques. *International Journal for Research in Applied Science and Engineering Technology*, v.10 n.3, p. 322-330, 2020. Disponível em:

https://www.researchgate.net/publication/359629890_A_Survey_on_Stock_Market_Price_Prediction_System_using_Machine_Learning_Techniques. Acesso em 14 ago. 2024.

KAMIJO, K. e TANIGAWA, T. Stock price pattern recognition-a recurrent neural network approach. *IJCNN International Joint Conference* v. 1 p. 215–221 2012. Disponível em: <https://ieeexplore.ieee.org/document/5726532>. Acesso em 14 ago. 2024.

KIM, Kyoung-jae. Financial time series forecasting using support vector machines. *Neurocomputing*. v. 55, n. 1-2, p 307-319, 2003. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0925231203003722>. Acesso em 15 ago. 2024.

LERTWACHARA K. Selecting stocks using a genetic algorithm: a case of real estate investment trusts (REITs). *Kasetsart Journal*. v. 28, n.28, p 106-116, 2007. Disponível em:

https://www.researchgate.net/publication/228884568_Selecting_stocks_using_a_genetic_algorithm_A_case_of_Real_Estate_Investment_Trusts_REITs. Acesso em 14 ago. 2024.

LEVIN, Andrew T.; LIN, Chien-Fu; CHU, Chia-Shang J. Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics*. v. 108, n.1, p 1-24, 2002. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0304407601000987>. Acesso em 11 fev. 2025.

LING D.C, NARANJO A. Dedicated REIT mutual fund flows and REIT performance. *J Real Estate Finance Econ*. v. 4, p 409–433, 2006 Disponível em:

<https://link.springer.com/article/10.1007/s11146-006-6960-y>. Acesso em 14 ago. 2024.

MACHADO, André Augusto de Almeida; CORREA, Geraldo. Avaliação de modelos de previsão dos valores das ações no mercado financeiro usando aprendizado de máquina. *Revista de Tecnologias*. v. 15, n. 2, 2022. Disponível em:

<https://www.fatecourinhos.edu.br/retec/index.php/retec/article/view/440>. Acesso em 15 ago. 2024.

MISHKIN, S. Frederic; SERLETIS, Apostolos. **The economics of money, banking, and financial Markets**. Pearson Canada. 4^oed. 2011. Disponível em: https://bibliotheque.pssfp.net/livres/THE_ECONOMICS_OF_MONEYS_BANKING_AND_FINANCIAL_MARKETS.pdf. Acesso em 14 ago. 2024

SANTOS, Murilo Alves; SOUZA, Daniel Henrique Silva; PENEDO, Antonio Sergio Torres; MARTINS, Elvis Silveira. Aplicação de redes neurais no Brasil: um estudo bibliométrico. **Relatos de Pesquisa**. v. 12, n. 2, 2016. Disponível em: [Aplicação de redes neurais no Brasil: um estudo bibliométrico | Biblionline \(ufpb.br\)](#). Acesso em 15 ago. 2024.

SEABOLD, Skipper; PERKTOLD, Josef. **Statsmodels: Econometric and Statistical Modeling with Python**, in: *9th Python in Science Conference*. 2010. Disponível em: <https://www.statsmodels.org/>. Acesso em 30 jan. 2025.

SILVEIRA, R. B. **Avaliação de modelos preditivos para o mercado de ações utilizando machine e deep learning**. Trabalho de Conclusão de Curso, Centro de Ciências Exatas e Engenharias - Universidade de Caxias do Sul-UCS, Caxias do Sul, 2021. Disponível em: <https://repositorio.ucs.br/xmlui/handle/11338/9714>. Acesso em 14 ago. 2024.

TAVARES, Ubiratan da Silva; Quadrelli, Giovane. Análise comparativa entre os modelos ARIMA e LSTM na previsão de curto prazo da demanda de potência ativa. **Revista de Engenharia e Tecnologia**. v. 14, n. 1, p 60-70, 2016. Disponível em: <https://revistas.uepg.br/index.php/ret/article/view/19752>. Acesso em 15 ago. 2024.

WOOLDRIDGE, Jeffrey M. *Introductory Econometrics, a Modern Approach*. **South-Western College Publishing, a division of Thomson Learning**, 2000. Disponível em: https://cbpbu.ac.in/userfiles/file/2020/STUDY_MAT/ECO/2.pdf. Acesso em 15 ago. 2024.

ZHANG, Wendi; LI, Bing; LIEW, Alan Wee-Chung; ROCA, Eduardo; SINGH, Tarlok. Predicting the returns of the US real estate investment trust market: evidence from the group method of data handling neural network. **Financial Innovation**. v. 9, n. 98, 2023. Disponível em: <https://jfin-swufe.springeropen.com/articles/10.1186/s40854-023-00486-2>. Acesso em 15 ago. 2024.