

## APLICAÇÃO DE APRENDIZADO DE MÁQUINA NA PREVISÃO DE PREÇOS DE ALUGUEL DE IMÓVEIS

Matheus Miranda  
Graduado em Ciências da Computação – Uni-FACEF  
matheusmiranda2809@gmail.com

Pedro Luiz Zuviollo  
Graduado em Ciências da Computação – Uni-FACEF  
pedrozuvi@gmail.com

Jaqueline Brigladori Pugliesi  
Docente do Uni-FACEF  
jbpugliesi@gmail.com

### Resumo

O mercado imobiliário desempenha um papel crucial na economia e na vida de várias pessoas envolvidas no processo comercial, sendo ele o responsável pelas ações de compra e aluguel de imóveis. Com isso em mente, é possível notar que a habilidade de prever com acurácia os preços destes imóveis é importante para os inquilinos envolvidos neste mercado. Diversos métodos de realizar essa previsão foram tentados no passado, mas com a constante evolução da tecnologia, uma nova possibilidade se revelou nos anos recentes: o uso da Inteligência Artificial (IA). Este trabalho utiliza algoritmos de IA para criar um modelo preciso de previsão de preços de aluguel de imóveis. Seu objetivo é fornecer aos inquilinos informações valiosas para decisões financeiras e identificar pontos de interesse. O estudo inclui uma estrutura teórica e procedimentos metodológicos com análise de dados. Para atingir esses objetivos, foi usada uma fundamentação teórica a partir do atual estado do mercado imobiliário juntamente aos conceitos e técnicas envolvidas na criação de um modelo de IA confiável. Também foram aplicados procedimentos metodológicos baseados na análise de dados do mercado imobiliário para o treino e finalização dos modelos de aprendizagem, que passaram posteriormente por uma avaliação para definir qual deles possui as previsões mais acuradas e confiáveis. Os resultados esperados deste estudo incluem a criação de um modelo de previsão de preços coerentes e confiáveis, além da descoberta das variáveis mais impactantes na definição desses valores.

**Palavras-chave:** Mercado imobiliário. Aprendizado de máquina. Modelo de previsão.

### Abstract

The real estate market plays a crucial role in the economy and in the lives of various individuals involved in the commercial process, as it is the one responsible for the actions of buying and renting properties. With that in mind, it is possible to note that the ability to accurately predict the prices of these properties is important for the tenants involved in this market. Several methods of making this prediction have been attempted in the past, but with the constant evolution of technology, a new possibility has emerged in recent years: the use of Artificial Intelligence (AI). This project uses AI algorithms to create an accurate model for predicting rental property prices. Its goal is

to provide valuable information to tenants for financial decisions and to identify points of interest. The study includes a theoretical framework and methodological procedures with data analysis. To achieve these objectives, a theoretical foundation was used based on the current state of the real estate market along with the concepts and techniques involved in creating a reliable AI model. Methodological procedures based on the analysis of real estate market data were also applied to train and finalize the learning models, which were subsequently evaluated to determine which one provides the most accurate and reliable predictions. The expected results of this study include the creation of a coherent and reliable price prediction model, as well as the discovery of the most impactful variables in defining these values.

**Keywords:** Real estate market. Machine learning. Prediction model.

## 1 Introdução

O mercado imobiliário desempenha um papel economicamente relevante, impactando compras, vendas e aluguéis. Valores acurados são significativos para os inquilinos envolvidos no mercado. Prever preços com precisão pode fornecer informações úteis para decisões estratégicas (COSTA, 2023).

Com o avanço da tecnologia e a disponibilidade de grandes volumes de dados, técnicas de aprendizado de máquina têm sido amplamente utilizadas por pesquisadores e profissionais do setor imobiliário para prever os custos de imóveis. Essas técnicas se baseiam em algoritmos capazes de analisar padrões e relações entre características dos imóveis, como localização, tamanho, número de quartos e banheiros, entre outros, e seus respectivos preços de aluguel ou de venda.

Este trabalho visa criar um projeto de aprendizado de máquina para prever os preços de aluguel de imóveis. Usando algoritmos de regressão, foi desenvolvido um modelo para fornecer estimativas confiáveis dos valores, beneficiando a compreensão do mercado e auxiliando nas negociações dos inquilinos. O estudo investiga a influência de variáveis nos custos, identificando as de maior impacto. Tais *insights* informam estratégias de precificação mais eficazes e baseadas em dados.

O trabalho empregou pré-processamento, seleção de variáveis, treinamento e avaliação de modelos de aprendizado de máquina. Utilizou de um conjunto de dados com informações físicas, localização e preços de imóveis. Algoritmos de regressão construíram um modelo preditivo. Também visou relatar o uso de aprendizado de máquina no mercado imobiliário, destacando a aplicação de regressão na previsão correta dos preços de aluguel.

## 2 O Mercado imobiliário

O mercado imobiliário é um setor econômico que envolve a compra, venda, locação e construção de imóveis. Ele é composto por diversas partes interessadas, como investidores, construtoras, corretores, proprietários de imóveis, inquilinos e compradores.

A dinâmica desse mercado é influenciada por diversos fatores, como a oferta e a demanda de imóveis, as condições econômicas e políticas, as tendências culturais e a evolução tecnológica, como referido por Barros (2021).

Os preços dos imóveis podem ser afetados por uma série de fatores, como a localização, o tamanho, a qualidade da construção, a infraestrutura e a oferta e demanda. Além disso, outros fatores, como as políticas de crédito e as taxas de juros, também podem influenciar os custos dos imóveis e a dinâmica do mercado (BARROS, 2021).

O mercado imobiliário é um setor importante para a economia, pois pode gerar empregos, aumentar o valor das propriedades, fornecer moradia para as pessoas e contribuir para o desenvolvimento de uma região. No entanto, também pode apresentar riscos e desafios, como a flutuação dos custos, a concorrência, a burocracia e a gestão de imóveis. É um mercado que pode ser muito lucrativo para investidores que buscam ganhos a longo prazo, como a valorização do imóvel e a geração de renda por meio do aluguel.

Apesar dos desafios, o mercado imobiliário continua sendo uma importante área de investimento e negócios em todo o mundo, e tem um papel fundamental no desenvolvimento econômico de uma região (COSTA, 2023).

## 2.1 Considerações ao Alugar um Imóvel

O processo de aluguel de imóveis envolve várias etapas complexas que requerem atenção meticulosa, visando evitar problemas futuros. Em meio aos desafios que podem surgir durante esse processo, destacam-se as pendências, conforme mencionado por especialistas como Fernandes (2019). Essas pendências podem englobar:

- **Documentação pendente:** a ausência de documentos essenciais, como contratos de locação anteriores ou a documentação do imóvel, pode retardar o processo de aluguel. É vital verificar se toda a documentação necessária está completa antes de seguir adiante.
- **Dívidas pendentes:** o imóvel pode ter dívidas pendentes, como taxas de condomínio em atraso ou impostos não quitados. Além disso, podem existir ônus, como hipotecas ou penhoras. É crucial assegurar que todas as dívidas e encargos estejam regularizados antes de prosseguir com o aluguel.
- **Questões estruturais:** problemas estruturais, como fissuras, vazamentos ou falhas elétricas, podem afetar a qualidade do imóvel. Realizar uma inspeção técnica antes de fechar o contrato de locação é importante para evitar surpresas desagradáveis.
- **Questões legais:** o imóvel pode estar envolvido em processos judiciais, como ações de despejo ou litígios. É fundamental verificar se o imóvel está livre de pendências judiciais antes de concretizar o aluguel.
- **Defeitos ocultos:** problemas ocultos, como questões de encanamento ou infestações, podem não ser visíveis durante a inspeção inicial. É prudente estar ciente desses problemas antes de finalizar o contrato de aluguel e, se possível, incluir cláusulas de manutenção no acordo.

Nesse contexto, é imprescindível que os potenciais locatários estejam conscientes das possíveis pendências e tomem medidas apropriadas para evitá-las, assegurando um processo de aluguel tranquilo e seguro.

## 2.2 Aspectos Financeiros no Aluguel de Imóveis

Quando se trata do aluguel de imóveis, a análise minuciosa dos aspectos financeiros é uma etapa crucial para uma tomada de decisão. Mais do que apenas o valor mensal do aluguel, diversos fatores financeiros desempenham um papel fundamental na determinação da viabilidade e sustentabilidade dessa escolha, como mencionado por Fernandes (2019). Estes podendo ser:

- **Custo mensal:** antes de tudo, é crucial calcular como o valor do aluguel se encaixa no orçamento mensal. Isso inclui considerar a renda disponível após deduzir todas as despesas fixas, como contas, alimentação e transporte.
- **Taxas adicionais:** além do aluguel, muitos imóveis têm taxas mensais de condomínio e o IPTU anual. Essas despesas adicionais precisam ser consideradas para determinar o custo total da locação.
- **Reajustes anuais:** verificar se o contrato prevê aumentos anuais no valor do aluguel e entender qual é o índice utilizado para o reajuste. Isso é importante para se preparar para potenciais aumentos de custo ao longo do tempo.
- **Custos de manutenção:** a manutenção e reparos podem ser necessários ao longo do período de locação. Estimar esses custos é essencial para evitar surpresas financeiras, principalmente se o imóvel não estiver em perfeitas condições.
- **Localização e valorização:** a localização do imóvel desempenha um papel significativo em sua valorização ao longo do tempo. Uma área em crescimento pode aumentar o valor do imóvel, proporcionando um benefício financeiro a longo prazo.
- **Termos do contrato:** ler atentamente o contrato é fundamental para compreender os termos de pagamento, cláusulas de rescisão antecipada e eventuais penalidades. Isso ajuda a evitar custos não previstos.
- **Comparação de ofertas:** analisar diferentes opções de aluguel permite uma comparação direta entre os custos e benefícios de cada imóvel, ajudando a escolher a opção mais vantajosa para o orçamento.
- **Negociação:** em alguns casos, é possível negociar o valor do aluguel ou outros aspectos do contrato. Estar aberto à negociação pode resultar em condições financeiras mais favoráveis.
- **Estabilidade financeira:** considerar a estabilidade da renda é fundamental. Certificar-se de que o aluguel não compromete outras responsabilidades financeiras ou objetivos futuros é primordial.
- **Benefícios fiscais:** é aconselhável consultar um especialista para entender possíveis vantagens fiscais, como deduções de imposto de renda relacionadas aos custos do aluguel.
- **Tempo de estadia:** avaliar por quanto tempo planeja permanecer no imóvel ajuda a determinar se os custos iniciais são justificados em relação ao período de permanência.

Ao considerar esses aspectos financeiros de maneira abrangente, os potenciais inquilinos estarão mais preparados para tomar uma decisão consciente e favorável às suas circunstâncias financeiras e objetivos de longo prazo.

### 3 Aprendizado de Máquina

O termo aprendizado de máquina, também conhecido como *machine learning*, teve o seu primeiro uso no ano de 1959 pelo engenheiro Arthur Samuel que, com base nos trabalhos do pai da computação, Alan Turing, da mesma década, envolvendo experimentos para testar a capacidade das máquinas de raciocinarem, conseguiu desenvolver o primeiro programa com a habilidade de aprender um jogo de damas no qual o sistema melhora o seu desempenho a cada partida jogada (NEOWAY, 2020).

Vários cientistas da computação propuseram diferentes definições para essa tecnologia capaz de fazer com que computadores possam aprender sem serem explicitamente programados, contudo, a definição de Samuel continua até hoje a ser uma das mais fáceis de se entender (ALMEIDA, 2020).

Após o primeiro uso do termo por Samuel, a área de aprendizado de máquina continuou a ser aprimorada e a evoluir exponencialmente, com computadores cada vez mais avançados sendo desenvolvidos a cada ano. Em razão disso, o aprendizado de máquina continua até hoje a ser um campo de grande importância e influência em todo o mundo, sendo aplicado em diferentes áreas para a resolução dos mais diversos problemas.

#### 3.1 O que é aprendizado de máquina

O aprendizado de máquina é uma das maiores áreas dentro da inteligência artificial (IA), consistindo essencialmente na ideia e no desenvolvimento de máquinas que podem aprender, de forma parcial ou autônoma, a realizar tarefas específicas a partir da sua própria experiência e da análise e interpretação de grandes volumes de dados (NEOWAY, 2020).

São computadores capazes de pensar e aprender a partir dos seus erros de forma equivalente à dos humanos e graças ao seu aspecto iterativo, ou seja, à sua habilidade de se tornar cada vez mais eficiente com cada uso sucessivo, podem tomar decisões e realizar diversas tarefas por si mesmos com mínima intervenção humana (FACELI *et al.*, 2021).

O funcionamento do aprendizado de máquina envolve o uso de diversos algoritmos responsáveis por diferentes comandos que, quando combinados, permitem com que o computador consiga adaptar e tomar decisões baseadas nas situações e informações que encontrar. Esse fato está ligado ao processo de aprendizagem do sistema, que se inicia com a coleta de dados históricos que servirão como dados de treinamento para a máquina. Após ser fornecido um conjunto de dados históricos já classificados ao sistema, o algoritmo começará o processo de aprendizagem a partir desses dados e, no final, o usuário terá um modelo treinado. Esse novo modelo poderá então ser alimentado com novos dados e, com base no que aprendeu durante o seu treinamento, conseguirá fazer previsões e chegar a novas conclusões (DIDÁTICA TECH, s.d.).

Existem quatro tipos de aprendizado de máquina, com cada algoritmo sendo melhor em uma situação diferente, como citado por Almeida (2020). Estes quatro métodos são:

- **Aprendizado supervisionado:** esse método consiste em fornecer ao sistema um conjunto de dados em que a resposta correta já está clara. A máquina aprenderá a diferenciar os dados corretos dos incorretos baseado nas características dos dados que já sabe que estão corretos. Por exemplo, para ensinar um sistema de aprendizagem de máquina a diferenciar um animal de outros com este método, basta fornecer a ele fotos desse animal com o seu nome e fotos de outros animais com outro título como “animal errado”. O sistema então irá aprender por meio de seus algoritmos a diferenciar o animal selecionado dos outros com base nas suas características presentes nas fotos. Dessa forma, quando o sistema for alimentado com um novo conjunto de dados sobre um animal que ele nunca viu antes, ele conseguirá identificar esse animal por si mesmo.
- **Aprendizado não supervisionado:** esse método envolve fornecer ao sistema dados que não possuem rótulos, ou seja, que não possuem uma resposta certa específica. Como resultado, a máquina deverá analisar os dados e chegar às suas próprias conclusões. Por exemplo, se um sistema com este modelo de aprendizagem receber um conjunto de dados de *marketing*, é possível que ele encontre associações ou padrões entre esses dados que possam ajudar o seu usuário a tomar uma decisão para beneficiar a sua empresa.
- **Aprendizado semi-supervisionado:** esse método funciona como uma junção dos primeiros métodos e envolve fornecer ao sistema um conjunto de dados no qual parte das informações são corretas e incorretas e a outra parte não possui uma resposta. A máquina deverá então utilizar os dados corretos e incorretos para aprender a descobrir quais dos desconhecidos sem resposta são corretos. As melhores suposições do sistema são em seguida devolvidas ao mesmo para o ajudar a melhorar o seu desempenho.
- **Aprendizado por reforço:** o modelo difere dos outros três por não utilizar dados prévios. Nesse método, a máquina aprende por tentativa e erro, recebendo recompensas direcionadas. A máquina interage com um ambiente, buscando adaptação por meio de múltiplas tentativas. Durante esse processo, recebe recompensas positivas quando suas ações atingem o objetivo, e negativas quando não produzem os resultados desejados. Essas recompensas são *feedbacks* que guiam o ajuste de estratégias, incentivando a identificação das ações corretas para maximizar o objetivo. Esse método é usado em áreas como jogos, robótica e otimização de processos, permitindo à máquina aprender a tomar decisões eficazes com base na experiência acumulada.

### 3.2 A importância do aprendizado de máquina e as suas aplicações

Com a habilidade de se desenvolver sistemas autômatos que podem aprender a realizar diversas tarefas específicas por si mesmos a partir de apenas um conjunto de dados inicial fornecido a eles, a área do aprendizado de máquina viu um aumento constante de usos durante os anos, com cada vez mais formas de se utilizar esses sistemas surgindo conforme a tecnologia continuou a avançar, como mencionado por Faceli *et al.* (2021). O conceito do aprendizado de máquina surgiu do reconhecimento de padrões, como referido por Neoway (2020), com essa sendo uma

aplicação desses sistemas que ainda é bastante usada, contudo, vários outros usos foram surgindo com o tempo, com alguns desses podendo ser vistos na maioria das empresas modernas que trabalham com grandes volumes de dados. Serviços financeiros, como bancos e operadoras de cartões de créditos, por exemplo, necessitam bastante de sistemas de detecção de fraude para identificar atividades ou transações suspeitas na conta de um cliente, permitindo que clientes sujeitos a fraudes sejam contatados para confirmar a atividade e possivelmente prevenir a fraude (FACELI *et al.*, 2021).

Operadoras de saúde, como grandes instituições médicas e empresas do setor de seguro saúde, também utilizam o aprendizado de máquina na forma de dispositivos e sensores que permitem com que os médicos consigam acessar e acompanhar, em tempo real, dados de seus pacientes. Isso pode os ajudar a detectarem doenças ou alertas de possíveis doenças nos pacientes, fazendo com que eles possam começar o seu tratamento mais cedo e, possivelmente, salvando as suas vidas (SAS, s.d.).

Veículos autônomos são outro elemento visto no nosso dia a dia que faz uso de aprendizado de máquina. O desenvolvimento de algoritmos de aprendizado de máquina é essencial para a condução autônoma desses veículos, que exigem identificação e tomadas de decisões com base em dados de sensores e câmeras (NEOWAY, 2020).

Até mesmo o governo necessita do uso de algoritmos de aprendizado de máquina devido à grande quantidade de dados que podem ser minerados para prover a essas agências governamentais diversas informações úteis que poderão ser usadas para a tomada de decisões importantes (SAS, s.d.).

Por fim, uma grande aplicação do aprendizado de máquina nos tempos modernos vem na forma de previsões. Os algoritmos de aprendizado de máquina podem ser usados para analisar dados históricos e, por exemplo, fazer previsões das demandas de certos produtos de forma rápida e eficiente, como citado por Faceli *et al.* (2021). Este uso tem sido de alta importância para indústrias e empresas. Outros exemplos dessa aplicação incluem previsões de preços e até mesmo de desempenhos. Estas predições são de grande valor durante uma tomada de decisão, seja ela sobre uma deliberação empresarial importante ou até mesmo sobre como o usuário deveria gastar o seu dinheiro.

Em geral, a área do aprendizado de máquina tem tido muitos usos em anos recentes, com exemplos adicionais aos já mencionados. Isso inclui práticas como análise e coleta de dados, sites de busca, construção de robôs, leituras biométricas e até reconhecimento de voz, como mencionado por Neoway (2020). O auxílio na tomada de decisões, o aumento da eficiência (devido à habilidade de analisar bilhões de dados em segundos) e a ajuda em identificar as melhores oportunidades e decisões estratégicas tornam essa subárea da IA uma das mais importantes para o mercado de trabalho atual.

#### **4 Ferramentas e metodologias utilizadas**

Antes da realização de um projeto, é necessário definir não só a ideia do projeto, mas também quais as ferramentas e métodos serão utilizadas durante o seu desenvolvimento. Nesta seção, são discutidas as ferramentas e métodos escolhidos para a realização deste projeto.

## 4.1 A linguagem Python

A linguagem de programação adotada é o Python, uma criação de Guido van Rossum no ano de 1991. Python se destaca como uma linguagem de alto nível, interpretada e orientada a objetos. Seu prestígio aumentou notavelmente, especialmente na área da ciência de dados, devido à sua facilidade de uso, clareza de código e uma ampla gama de bibliotecas disponíveis. Na esfera da ciência de dados, Python é frequentemente empregado para tarefas como manipulação, processamento e análise de dados, visualização e aprendizado de máquina (PYTHON SOFTWARE FOUNDATION, s.d.).

Um dos pilares fundamentais no âmbito da ciência de dados em Python é a Pandas. Pandas é uma biblioteca *open-source* que oferece estruturas de dados para a análise de informações. Construída sobre a base da NumPy, ela se destaca por ser uma ferramenta essencial na manipulação de dados. Sua aplicabilidade pode ser observada na abordagem de dados tabulares, como planilhas CSV (Comma-Separated Values) e Excel, dispondo de funcionalidades para leitura, escrita e transformação desses dados. As características da biblioteca Pandas também englobam a capacidade de limpeza de dados, tratamento de valores faltantes e aplicação de transformações (PANDAS, 2023). Adicionalmente, proporciona funcionalidades avançadas, tais como agregação de dados, operações em grupos e fusão de informações. Comumente, o Pandas é empregado de forma conjunta com outras bibliotecas em Python, a exemplo do Matplotlib e do Seaborn, para a visualização de dados.

A visualização de dados desempenha um papel crucial tanto na análise de informações quanto na ciência de dados de maneira geral. Por meio da representação visual dos dados, é possível discernir padrões, tendências e *insights* que poderiam passar despercebidos por outros meios. Nesse contexto, o Matplotlib, uma biblioteca de visualização de dados em Python, se destaca por sua versatilidade (MATPLOTLIB, 2023). O Matplotlib oferece uma vasta gama de tipos de gráficos, incluindo gráficos de linhas, de barras e de dispersão. Através de muita customização, é viável ajustar elementos como cores, legendas e eixos. A sua natureza interativa contribui para sua popularidade nessa aplicação, sendo especialmente útil na exibição de dados tabulares e na representação de tendências temporais.

Outra ferramenta relevante na visualização de dados em Python é o Seaborn. Construído sobre o Matplotlib, o Seaborn oferece gráficos mais avançados e customizáveis, abrangendo opções como gráficos de densidade, gráficos de dispersão com ajuste de regressão e gráficos de distribuição (SEABORN, 2023). Esse conjunto de recursos se mostra particularmente eficaz na visualização de dados categóricos, como é o caso de histogramas e gráficos de barras empilhadas. Além disso, o Seaborn disponibiliza suporte para a criação de mapas de calor (*heatmaps*), que revelam correlações entre diferentes variáveis. Uma característica marcante do Seaborn é a sua extensa customização, que oferece um leque amplo de opções para ajustar a estética dos gráficos produzidos.

No campo da ciência de dados em Python, uma outra biblioteca de destaque é o Scikit-learn. Trata-se de uma biblioteca *open-source* voltada para aprendizado de máquina, fornecendo ferramentas para construir modelos de aprendizado de máquina em Python. Dentro do Scikit-learn estão disponíveis variados algoritmos de aprendizado de máquina, incluindo regressão linear, regressão

logística, árvores de decisão e redes neurais (SCIKIT-LEARN, s.d.). Adicionalmente, essa biblioteca oferece funcionalidades para validação de modelos, seleção de atributos e pré-processamento de dados. O Scikit-learn possui ampla utilização em projetos envolvendo aprendizado de máquina em Python, sendo considerado uma das principais bibliotecas para a área de ciência de dados. A combinação do Pandas com o Scikit-learn se revela valiosa para a criação de modelos precisos de aprendizado de máquina, incluindo a previsão de preços de imóveis.

#### 4.2. Os algoritmos de aprendizado de máquina

Entre os algoritmos de aprendizado de máquina mais comuns, encontram-se a regressão linear, as árvores de decisão, as redes neurais e os métodos de *ensemble*, como o Random Forest e o Gradient Boosting. Além disso, há algoritmos de agrupamento, como o *k-means*, e algoritmos de classificação, como Support Vector Machine (SVM) e os classificadores de Bayes. Cada algoritmo possui características específicas e é aplicado em situações distintas, dependendo do problema e das características dos dados (FACELI *et al.*, 2021).

A regressão linear é usada para modelar a relação entre uma variável de saída contínua e um conjunto de variáveis de entrada. Ela pressupõe uma relação linear e ajusta os coeficientes da equação para melhor se adequarem aos dados. Essa técnica é comumente empregada para previsões e estimativas numéricas, como prever os valores dos imóveis baseados em suas características (MÜLLER e GUIDO, 2016).

As árvores de decisão são estruturas em forma de árvore, em que cada nó representa uma decisão ou teste em uma característica específica. Dividem o conjunto de dados com base em critérios definidos, gerando ramos que levam a diferentes resultados. São altamente interpretáveis e podem ser usadas para classificação ou regressão, tornando acessíveis as regras e padrões identificados (MÜLLER e GUIDO, 2016).

As redes neurais, inspiradas no cérebro humano, consistem em camadas de neurônios interconectados. Cada neurônio recebe entradas ponderadas, aplica uma função de ativação e propaga o resultado. São capazes de aprender representações complexas e relações não lineares, sendo apropriadas para problemas complexos e de alta dimensionalidade (PURESTORAGE, s.d.).

Métodos de *ensemble* combinam diversos modelos de aprendizado para aumentar a precisão e robustez das previsões. O Random Forest constrói várias árvores de decisão independentes e as combina, resultando em previsões mais sólidas. O Gradient Boosting mescla árvores sequencialmente, melhorando gradualmente as previsões (FACELI *et al.*, 2021).

Em adição, os algoritmos de agrupamento, como o *k-means*, identificam grupos nos dados com base em características similares. São usados em análises exploratórias para identificar padrões e grupos (PURESTORAGE, s.d.).

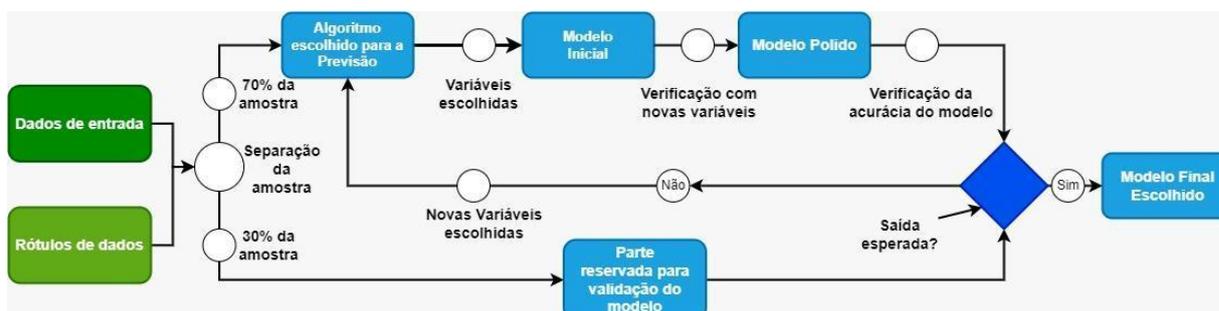
Os algoritmos de classificação, como SVM e classificadores de Bayes, atribuem rótulos a dados. No contexto de preços de imóveis, podem classificar em faixas de preço ou determinar se um imóvel está acima ou abaixo de um valor referência (PURESTORAGE, s.d.).

A escolha do algoritmo depende do problema, dos dados e dos objetivos. Cada algoritmo tem vantagens e limitações, exigindo análise criteriosa para a seleção adequada.

### 4.3 Validação de um modelo

A validação de um modelo de aprendizado de máquina é importante para assegurar sua eficácia e habilidade de generalização. Como representado de forma visual e detalhada na Figura 1, esse processo começa com a separação da amostra dos dados. Uma parte da amostra é reservada para a escolha do algoritmo de aprendizado que será utilizado para criar o modelo. Isso envolve a experimentação com diferentes algoritmos, ajustando seus parâmetros e avaliando seu desempenho em termos de métricas específicas, como precisão, sensibilidade e especificidade. Após a seleção do algoritmo, a próxima etapa crucial é a escolha das variáveis que serão utilizadas para construir o modelo inicial e depois o otimizar. Uma vez que o modelo tenha sido criado com essas variáveis, ele é validado usando a outra parte das amostras. Isso garante que o modelo seja testado em dados independentes e fornece uma estimativa realista de seu desempenho em situações do mundo real (MIRANDA, 2016).

Figura 1 - Fluxograma do processo usado para validação do modelo



Fonte: os autores

## 5 Planejamento do Modelo

Para iniciar a criação de um ambiente voltado à análise de dados e à construção de modelos de regressão linear, SVM e árvore de decisão, buscamos na plataforma Kaggle um conjunto de dados adequado. Utilizamos a base de dados *brazilian\_houses\_to\_rent*, disponível no site Kaggle, como nossa fonte primária de informações. Esta base de dados contém 10.962 casas disponíveis para aluguel, cada uma com 13 classes diferentes, representando diversas características relevantes. Estas características incluem detalhes como a área do imóvel, a quantidade de salas, banheiros, estacionamentos e outros atributos relevantes. Vale ressaltar que, ao explorar esse conjunto de dados, observamos a possibilidade de presença de valores atípicos (*outliers*), que podem afetar nossas análises. Uma vez que os dados foram coletados por meio de um *WebCrawler*, algumas inconsistências podem estar presentes, incluindo possíveis erros nos registros. Além disso, é importante notar que entre as diferentes versões do conjunto de dados, podem ocorrer mudanças, como a adição ou exclusão de casas, bem como a inclusão de cidades adicionais. Preparamos cuidadosamente o ambiente de trabalho para permitir uma manipulação

precisa desses dados. A escolha do Kaggle como plataforma de referência proporcionou-nos uma rica e valiosa fonte de informações para análise e desenvolvimento do nosso modelo (KAGGLE, 2020).

## 5.1 Pré-processamento

Esse estágio inicial envolve uma série de etapas essenciais para garantir a qualidade, a consistência e a relevância dos dados antes de prosseguir com qualquer análise mais profunda. Questões cruciais do pré-processamento incluem a detecção e tratamento de valores ausentes, a padronização ou normalização de escalas para evitar viés em algoritmos sensíveis à magnitude, a codificação adequada de variáveis categóricas, como a variável "*animal*" que especifica se animais são aceitos no imóvel, e "*furniture*" que indica se o imóvel é mobiliado ou não. Além disso, a variável "*city*" que corresponde ao número da cidade onde o imóvel está localizado, também requer atenção especial. Outras variáveis incluem "*area*" representando a área do imóvel, "*rooms*" para a quantidade de salas, "*bathroom*" para o número de banheiros, "*parking spaces*" para o número de estacionamentos disponíveis, "*floor*" que representa a quantidade de andares do imóvel, "*hoa*" (*Homeowners Association*) que é a Associação de Proprietários de Casas, "*rent amount*" que corresponde ao preço do aluguel, "*property tax*" um imposto calculado sobre bens imóveis, "*fire insurance*" que é o seguro contra incêndio, e "*total*" que representa o preço total do imóvel, sendo uma soma dos valores das variáveis "*hoa*" "*rent amount*" "*property tax*" e "*fire insurance*". Todos esses fatores desempenham um papel fundamental na análise dos dados. O pré-processamento também envolve a remoção de *outliers*, ou seja, valores discrepantes que destoam do conjunto e que podem distorcer as análises e as conclusões. Isso poderá ser feito através de opções de tratamento que incluem a remoção dos *outliers* e a sua substituição por valores realistas, bem como a transformação e segmentação dos dados, levando em consideração as particularidades de cada variável.

### 5.1.1 Tratamento dos dados

Após escolhido o banco de dados apropriado, realizamos a verificação da dispersão dos dados e a presença ou não de *outliers*. Neste caso, estes objetos foram removidos do processo. Logo após, foi feita a análise da dispersão dos valores de cálculo para averiguar se seria necessária alguma normalização ou conversão para melhorar a precisão do futuro modelo. A Tabela 1 mostra as 10 primeiras linhas dos dados que foram utilizados para as análises.

**Tabela 1 – Amostra dos dados usados para as análises**

Unnamed: 0	city	area	rooms	bathroom	parking spaces	floor	animal	furniture	hoa	rent amount	property tax	fire insurance	total	
0	0	1	240	3	3	4	-	accept	furnished	R\$0	R\$8,000	R\$1,000	R\$121	R\$9,121
1	1	0	64	2	1	1	10	accept	not furnished	R\$540	R\$820	R\$122	R\$11	R\$1,493
2	2	1	443	5	5	4	3	accept	furnished	R\$4,172	R\$7,000	R\$1,417	R\$89	R\$12,680
3	3	1	73	2	2	1	12	accept	not furnished	R\$700	R\$1,250	R\$150	R\$16	R\$2,116
4	4	1	19	1	1	0	-	not accept	not furnished	R\$0	R\$1,200	R\$41	R\$16	R\$1,257
5	5	1	13	1	1	0	2	accept	not furnished	R\$0	R\$2,200	R\$42	R\$28	R\$2,270
6	6	1	55	1	1	1	2	accept	furnished	R\$1,950	R\$5,000	R\$375	R\$64	R\$7,389
7	7	1	55	2	2	1	2	accept	not furnished	R\$600	R\$1,300	R\$67	R\$17	R\$1,984
8	8	1	82	3	1	1	3	accept	furnished	R\$828	R\$2,000	R\$109	R\$26	R\$2,963
9	9	1	32	1	1	1	16	accept	not furnished	R\$533	R\$2,900	R\$0	R\$37	R\$3,470

Fonte: os autores

Após a análise dos atributos, optamos pela Remoção de Coluna Desnecessária, a coluna Unnamed: 0, que foi descartada do *DataFrame*, como pode ser visto na Tabela 2. Isso é feito devido ao fato de que é um índice irrelevante e desnecessário para a análise. O método *drop* foi usado para remover essa coluna.

**Tabela 2 - Dados após a remoção de atributos irrelevantes e substituição de valores ausentes**

city	area	rooms	bathroom	parking spaces	floor	animal	furniture	hoa	rent amount	property tax	fire insurance	total	
0	1	240	3	3	4	6	accept	furnished	0	8000	1000	121	9121
1	0	64	2	1	1	10	accept	not furnished	540	820	122	11	1493
2	1	443	5	5	4	3	accept	furnished	4172	7000	1417	89	12680
3	1	73	2	2	1	12	accept	not furnished	700	1250	150	16	2116
4	1	19	1	1	0	6	not accept	not furnished	0	1200	41	16	1257
5	1	13	1	1	0	2	accept	not furnished	0	2200	42	28	2270
6	1	55	1	1	1	2	accept	furnished	1950	5000	375	64	7389
7	1	55	2	2	1	2	accept	not furnished	600	1300	67	17	1984
8	1	82	3	1	1	3	accept	furnished	828	2000	109	26	2963
9	1	32	1	1	1	16	accept	not furnished	533	2900	0	37	3470

Fonte: os autores

As colunas referentes a valores financeiros, como *hoa*, *rent amount*, *property tax*, *fire insurance* e *total* foram tratadas para remover caracteres não numéricos e transformá-las em valores numéricos. Isso é feito usando expressões regulares para substituir os caracteres não numéricos por *strings* vazias, deixando apenas os valores numéricos. Em seguida, os dados dessas colunas foram convertidos para o tipo numérico. Os valores ausentes na coluna *floor* foram substituídos pela média da coluna. Isso é feito para garantir que as informações de piso não sejam perdidas e que o conjunto de dados permaneça consistente.

Na próxima etapa, duas colunas categóricas, *animal* e *furniture*, foram transformadas utilizando a técnica de codificação de rótulo. Isso envolve a

transformação de categorias em valores numéricos, facilitando a utilização dessas variáveis em modelos de aprendizado de máquina. A classe *LabelEncoder* da biblioteca *sklearn* é usada para isso.

Depois de todos esses passos de pré-processamento, o *DataFrame* resultante é exibido, mostrando as primeiras 10 linhas após o tratamento dos dados, as quais podem ser vistas na Tabela 3. Cada coluna foi convertida para um formato apropriado e está pronta para ser utilizada em análises posteriores.

Em resumo, a análise realizada executa a leitura e limpeza do conjunto de dados de aluguel de casas, transformando-o em um formato mais apropriado para análises e modelagens futuras. Ele abrange desde a remoção de informações desnecessárias até a conversão de valores categóricos e numéricos para tornar os dados mais adequados para a aplicação de técnicas de aprendizagem de máquina.

**Tabela 3** - Dados após o processo de codificação de rótulo

	city	area	rooms	bathroom	parking spaces	floor	animal	furniture	hoa	rent amount	property tax	fire insurance	total
0	1	240	3	3	4	6	0	0	0	8000	1000	121	9121
1	0	64	2	1	1	10	0	1	540	820	122	11	1493
2	1	443	5	5	4	3	0	0	4172	7000	1417	89	12680
3	1	73	2	2	1	12	0	1	700	1250	150	16	2116
4	1	19	1	1	0	6	1	1	0	1200	41	16	1257
5	1	13	1	1	0	2	0	1	0	2200	42	28	2270
6	1	55	1	1	1	2	0	0	1950	5000	375	64	7389
7	1	55	2	2	1	2	0	1	600	1300	67	17	1984
8	1	82	3	1	1	3	0	0	828	2000	109	26	2963
9	1	32	1	1	1	16	0	1	533	2900	0	37	3470

**Fonte:** os autores

### 5.1.2 Escolha de Variáveis

Para garantir uma aderência mais considerável do modelo, é necessário escolher as variáveis mais representativas em relação ao valor do aluguel do imóvel. Para isso, nós escolhemos os atributos que mais influenciam no valor total, estas sendo *hoa*, *rent amount* e *fire insurance*. Para a verificação das variáveis, inicialmente foi retirada uma subamostra aleatória de 70% das observações para o treinamento do modelo. Essa subamostra foi composta por todas as variáveis descritas na Seção 5.1 e foi usada para realizar uma simulação de comportamento preditivo. Posteriormente, o resultado do modelo foi comparado com os 30% restantes da amostra, verificando qualquer divergência em relação aos valores reais.

Usando a metodologia descrita na Seção 5.1.1, efetuou-se testes com todas as variáveis fornecidas e, a partir destas, foram escolhidos os atributos apresentados na Tabela 4 para calibrar o modelo.

**Tabela 4** - Atributos escolhidos para o treinamento

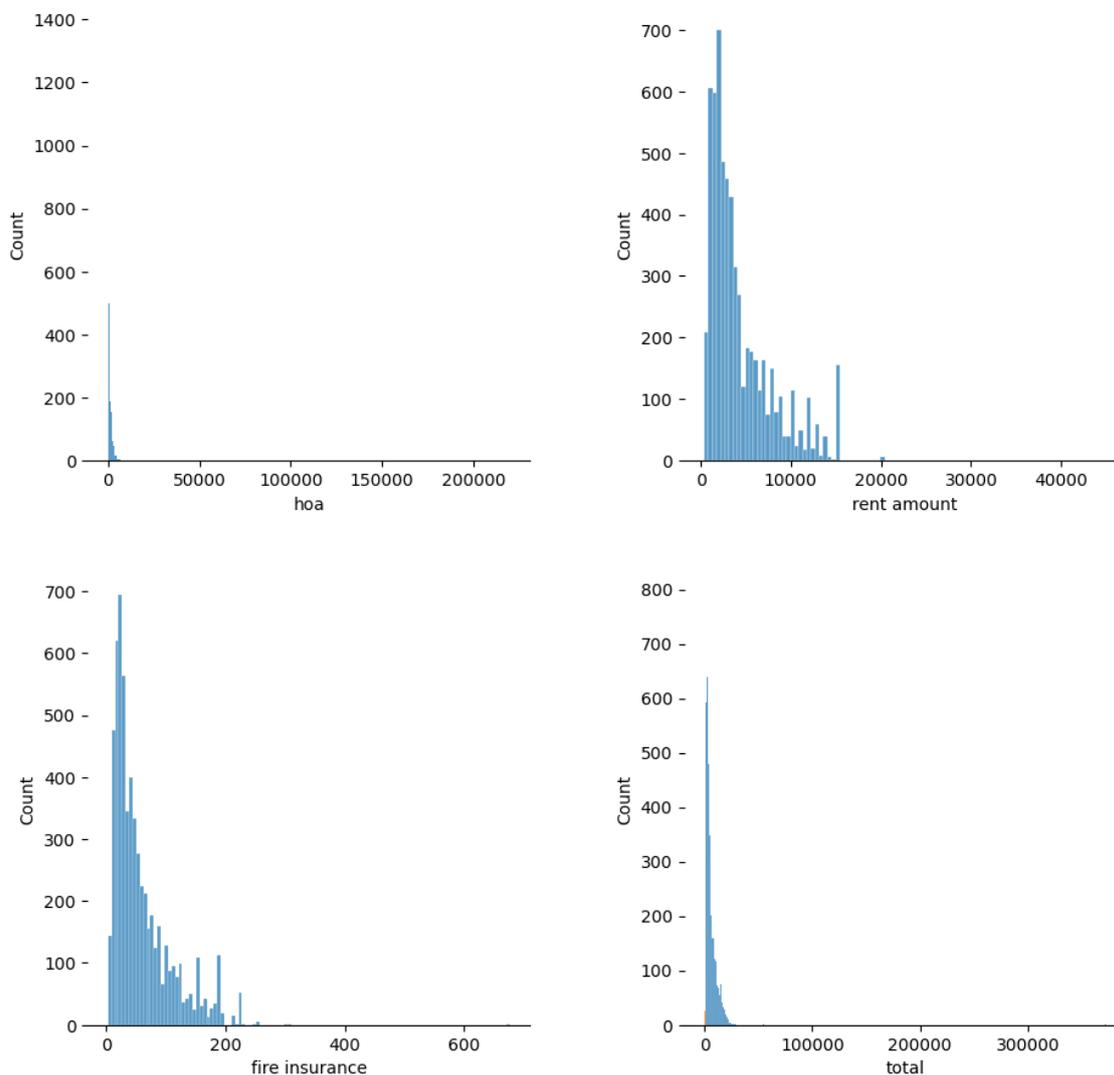
Variável	Tipo	
city	Discreta	Qualitativa
area	Contínua	Quantitativa
rooms	Discreta	Quantitativa
bathroom	Discreta	Quantitativa
parking spaces	Discreta	Quantitativa
floor	Discreta	Qualitativa
animal	Discreta	Qualitativa
furniture	Discreta	Qualitativa
hoa	Contínua	Quantitativa
rent amount	Contínua	Quantitativa
property tax	Contínua	Quantitativa
fire insurance	Contínua	Quantitativa
total	Contínua	Quantitativa

**Fonte:** os autores

## 5.2 Resultados obtidos

O processo de pré-processamento e análise exploratória revelou *insights* significativos sobre os dados do conjunto de habitações para aluguel, com foco nos atributos relacionados aos gastos de aluguel, sendo o *total* o atributo meta. Inicialmente, na etapa de distribuição dos dados, histogramas foram criados para as variáveis escolhidas, que incluem *hoa* (taxa de administração do condomínio), *rent amount* (valor do aluguel), *fire insurance* (seguro incêndio) e, em especial, o atributo *total*, devido a este englobar os preços das outras três variáveis junto com a variável *area*. Esses gráficos podem ser observados na Figura 2 e fornecem uma visão ampla da distribuição dos valores.

**Figura 2 - Histogramas para as variáveis selecionadas**

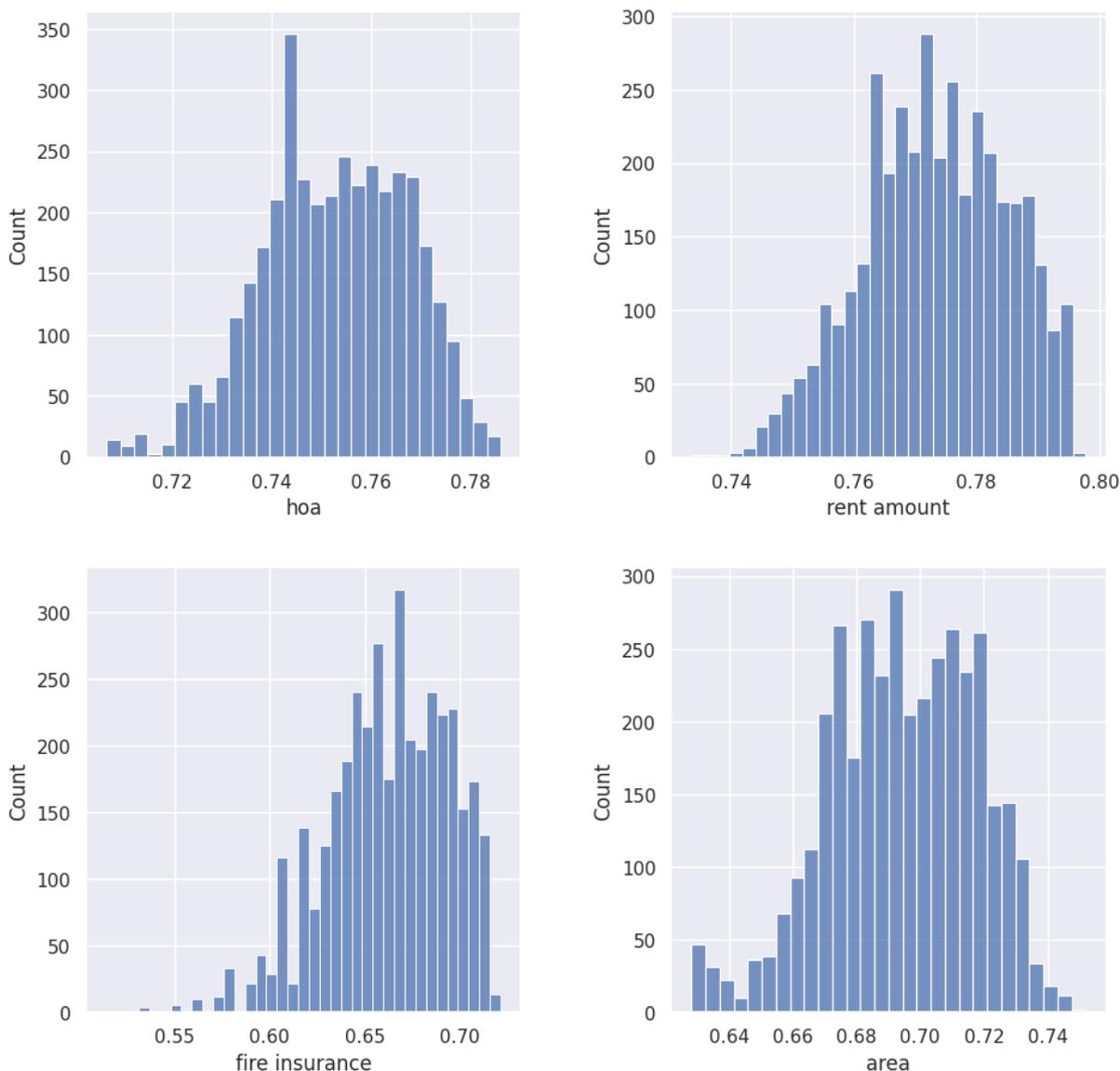


**Fonte:** os autores

Em seguida, após a aplicação de uma transformação logarítmica utilizando a função  $\log_{1p}$  para normalização, os histogramas foram recriados. Essa abordagem permitiu uma avaliação mais precisa da normalidade e redução de assimetrias nas distribuições, como pode ser visto na Figura 3.

Por fim, foi executada uma segmentação dos dados com base em faixas específicas para as quatro variáveis escolhidas. Como pode ser observado na Figura 4, os histogramas reavaliados revelaram uma maior concentração de dados dentro das faixas especificadas.

**Figura 3 - Histogramas recriados após transformação logarítmica**

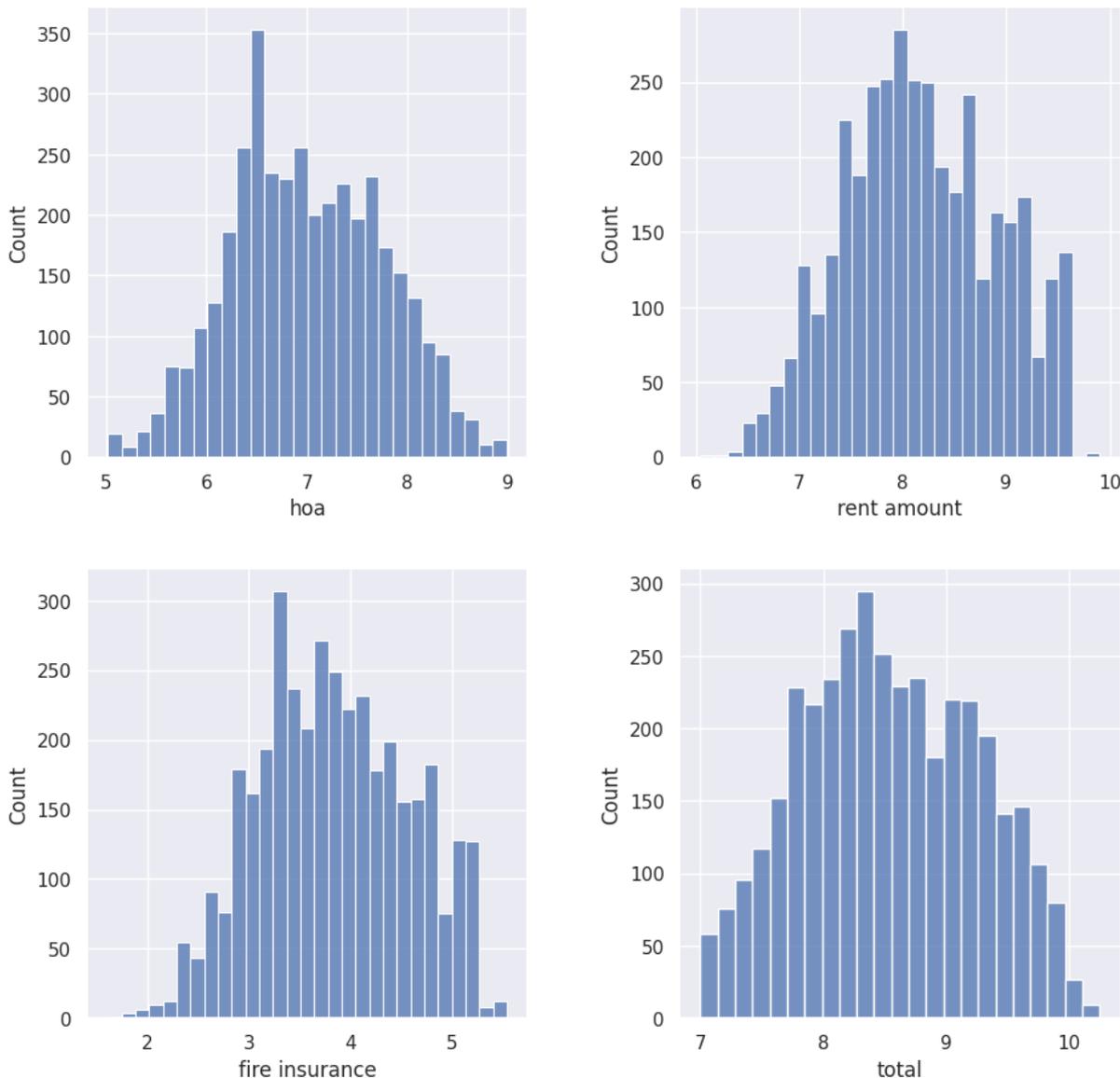


**Fonte:** os autores

Esta etapa crucial desempenhou um papel determinante na melhoria da qualidade e confiabilidade das informações analisadas. Ao definir intervalos específicos para as quatro variáveis, foi possível identificar e remover observações atípicas ou discrepantes que poderiam potencialmente distorcer as análises subsequentes. Essa abordagem foi essencial para assegurar que nossos histogramas fossem informativos e eficazes. A qualidade de um histograma é intrinsecamente ligada à sua capacidade de representar a distribuição dos dados de forma clara e precisa. Um histograma eficaz deve apresentar intervalos apropriados que capturem as variações nos dados, além de evidenciar tendências e padrões. Nossos histogramas, após a filtragem, exibiram uma maior concentração dos valores dentro das faixas especificadas, o que não apenas realçou a remoção das observações discrepantes, mas também destacou a qualidade da representação visual. Isso é fundamental para a interpretação de resultados precisos e a tomada de decisões informadas. Em resumo, a capacidade de um histograma em destacar características-

chave da distribuição dos dados é essencial para uma análise eficaz e significativa, fornecendo *insights* valiosos para a compreensão do fenômeno em questão.

**Figura 4 -** Histogramas após filtragem de dados



**Fonte:** os autores

### 5.3 Treinamento do Modelo

A próxima etapa do projeto envolveu o início do processo de treinamento do modelo. O treinamento de um modelo de regressão constitui um processo fundamental na análise de dados e na previsão de relações entre variáveis. O processo de treinamento envolveu a busca pelos coeficientes que minimizem a diferença entre os valores previstos pelo modelo e os valores reais presentes nos dados de treinamento.

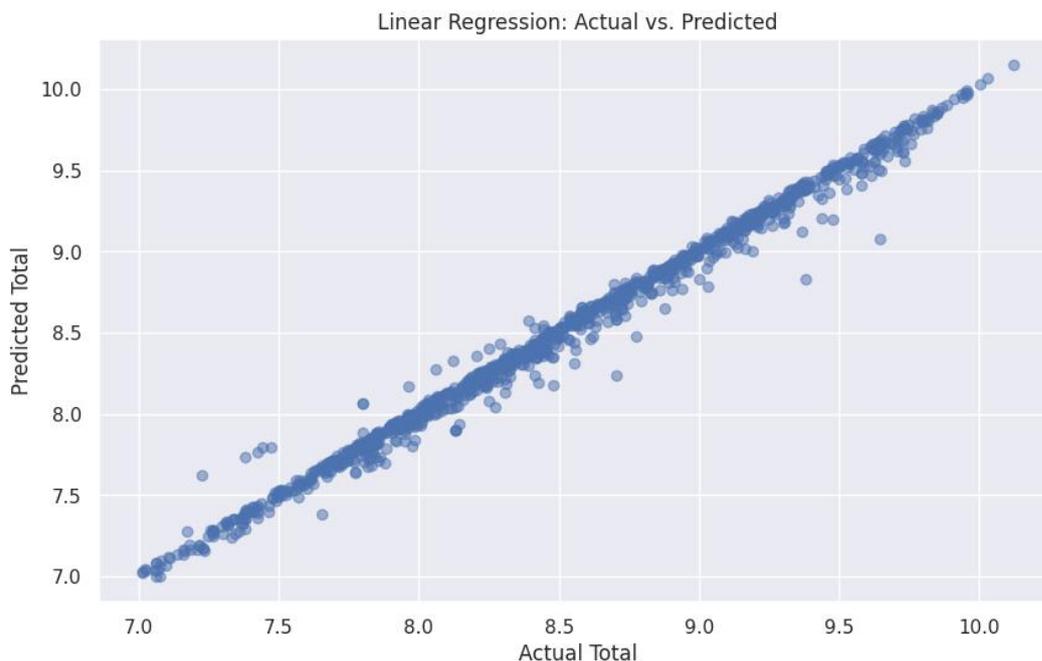
### 5.3.1 Regressão Linear

A princípio, foi conduzida uma análise de regressão linear utilizando todas as características disponíveis no conjunto de dados, com foco especial na variável de destino *total*. Esse procedimento visou identificar a relação entre as características independentes e o preço total do imóvel, permitindo que o modelo capturasse a complexidade das interações entre essas variáveis. Após a separação dos dados em conjuntos de treinamento e teste, o modelo de regressão foi avaliado quanto ao seu desempenho em ambos os conjuntos, incluindo a avaliação dos coeficientes do modelo e seu coeficiente de determinação  $R^2$  para o conjunto de treinamento. A equação da função de regressão linear encontrada foi a seguinte:

$$\text{total} = 0.82 + (0.00 * \text{feature\_1}) + (0.00 * \text{feature\_2}) + (-0.00 * \text{feature\_3}) + (0.01 * \text{feature\_4}) + (0.00 * \text{feature\_5}) + (0.00 * \text{feature\_6}) + (0.00 * \text{feature\_7}) + (-0.01 * \text{feature\_8}) + (0.24 * \text{feature\_9}) + (0.71 * \text{feature\_10}) + (0.05 * \text{feature\_11}) + (-0.01 * \text{feature\_12}).$$

Nesta equação, *total* representa o atributo meta que visamos prever, *feature\_1* até *feature\_12* são as características do conjunto de dados que foi usado como atributos independentes no modelo. O gráfico de acurácia desse algoritmo pode ser observado na Figura 5.

**Figura 5** - Gráfico da acurácia do modelo com regressão linear



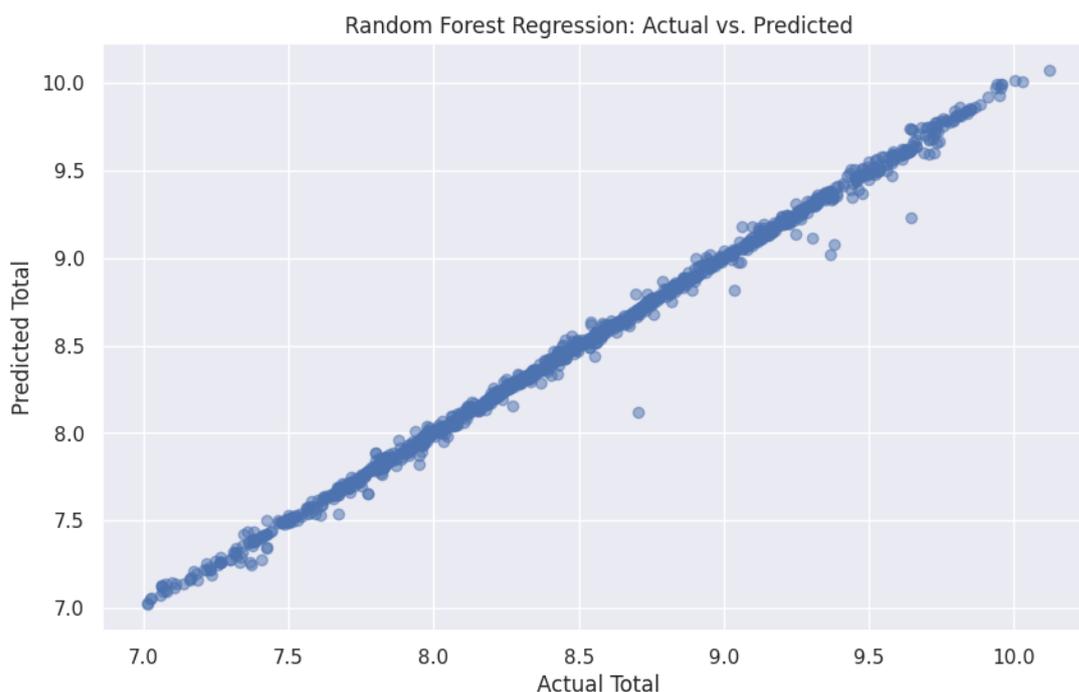
**Fonte:** os autores

### 5.3.2 Random Forests

Esta análise representou uma exploração abrangente ao incorporar um modelo de aprendizado de máquina diferente, como o Random Forest. Nessa etapa, o Random Forest Regressor foi implementado com 100 estimadores, permitindo assim a construção de um conjunto diversificado de árvores de decisão. Por meio deste *ensemble*, o modelo conseguiu capturar relações não lineares e interações complexas

presentes nos dados. O processo incluiu a capacitação do modelo com o conjunto de treinamento e a avaliação de seu desempenho por meio da métrica  $R^2$ , tanto nesse conjunto quanto no de teste. Ao comparar os resultados obtidos, que podem ser vistos na Figura 6, com o modelo de regressão linear (Figura 5), a análise ofereceu uma compreensão significativa do desempenho do modelo Random Forest. Além disso, também incluiu uma análise comparativa com o modelo de regressão linear, o que permitiu a seleção do modelo mais adequado às características dos dados e aos objetivos da análise.

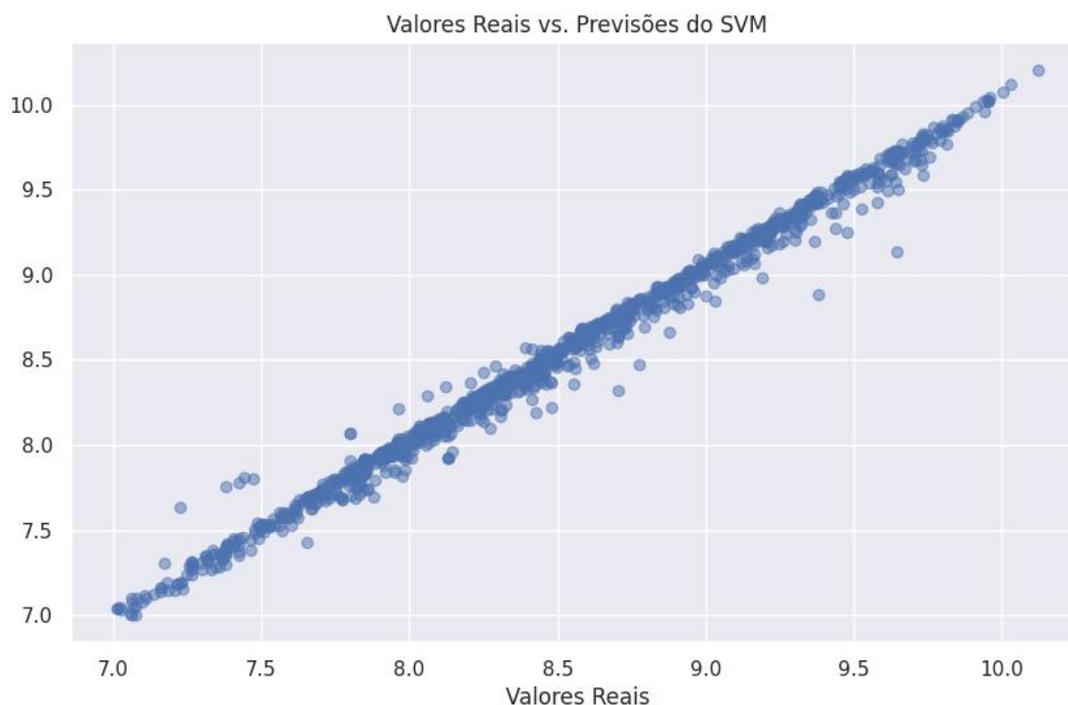
**Figura 6** - Gráfico da acurácia do modelo com random forest



Fonte: os autores

### 5.3.3 Support Vector Machine Regression

Essa análise introduziu a aplicação de um modelo de SVM, que é uma técnica de aprendizado de máquina que busca mapear os dados em um espaço de alta dimensão para encontrar um hiperplano que melhor se ajusta aos pontos. Nessa etapa, o modelo SVM com um *kernel* linear foi empregado, permitindo a captura de relações lineares entre as variáveis independentes e a variável de destino. Após ter capacitado o modelo com o conjunto de treinamento, a métrica  $R^2$  foi calculada tanto para ele quanto para o conjunto de teste, fornecendo uma avaliação significativa do desempenho do modelo. Essa etapa destacou a flexibilidade e adaptabilidade do SVM como uma abordagem alternativa à regressão linear e ao Random Forest, explorando o potencial de modelagem não linear dentro do contexto de previsão de preços de aluguel de imóveis. Estes resultados podem ser observados no gráfico de acurácia representado na Figura 7.

**Figura 7** - Gráfico da acurácia do modelo com SVM

**Fonte:** os autores

#### 5.4 Análise de Erro

Três algoritmos diferentes foram utilizados para realizar previsões e avaliar seu desempenho. A métrica aplicada foi o coeficiente de determinação  $R^2$ , que varia entre 0 e 1, com valores mais próximos de 1 indicando um melhor ajuste aos dados. O desvio padrão das previsões também foi calculado para medir a dispersão em relação aos valores reais. A validação cruzada foi realizada para verificar a capacidade de generalização dos modelos. Portanto, são apresentados os resultados dos modelos de Regressão Linear, Random Forest e SVM, com base nas métricas utilizadas para determinar sua acurácia e adequação na tarefa de previsão.

Para os três algoritmos de regressão foram explorados os resultados, sendo analisadas suas capacidades de ajuste aos dados, estabilidade nas previsões e capacidade de generalização, com base nos valores de  $R^2$  e desvios padrão no conjunto de treinamento e teste.

- **Regressão Linear:** para a Regressão Linear, o  $R^2$  no conjunto de treinamento foi de 0,9903, enquanto no conjunto de teste foi ainda mais alto, com 0,9910. Isso indica uma excelente capacidade do modelo de se ajustar aos dados. O desvio padrão nos valores previstos para a regressão linear também foi relativamente baixo, sugerindo que as previsões do modelo eram consistentes, com 0,0713 no treinamento e 0,0662 no teste. Esses resultados podem ser todos observados no Quadro 1.

- **Random Forests:** em relação ao modelo Random Forest, pode-se observar no Quadro 1 que os resultados foram ainda mais coerentes, com um  $R^2$  de 0,9996 no conjunto de treinamento e 0,9969 no conjunto de teste. O desvio padrão nos valores previstos para o Random Forest foi baixo tanto no treinamento (0,0141) quanto no teste (0,0391), indicando uma alta estabilidade nas previsões e reforçando a confiabilidade deste modelo.
- **Support Vector Machine Regression:** por fim, o modelo de SVM também apresentou resultados sólidos, como mostrado no Quadro 1, com um  $R^2$  de 0,9883 no conjunto de treinamento e 0,9895 no conjunto de teste. Os desvios padrão nos valores previstos para o SVM foram um pouco mais altos que os modelos anteriores, mas ainda aceitáveis, com 0,0734 no treinamento e 0,0688 no teste, indicando uma boa capacidade de generalização do modelo.

**Quadro 1** - Análise do erro dos modelos de regressão

	<b>R<sup>2</sup> - Treino</b>	<b>R<sup>2</sup> - Teste</b>	<b>Desvio Padrão - Treino</b>	<b>Desvio Padrão - Teste</b>
<b>Regressão Linear</b>	0,9903	0,9910	0,0713	0,0662
<b>Random Forest</b>	0,9996	0,9969	0,0141	0,0391
<b>SVM</b>	0,9883	0,9895	0,0734	0,0688

**Fonte:** os autores

Em conclusão, os modelos mostraram resultados satisfatórios, com alta precisão e baixa variabilidade nos valores previstos, sugerindo que são opções robustas para previsões de preços futuros com base nos dados disponíveis. Contudo, ainda existe espaço para aprimoramentos futuros do projeto e sua inclusão no mercado.

## 6 Considerações Finais

Neste artigo, exploramos a aplicação do aprendizado de máquina como uma ferramenta promissora na previsão de preços de aluguel de imóveis para inquilinos. Por meio da análise de um conjunto de dados, nosso objetivo foi compreender e modelar as complexas relações entre as características dos imóveis e seus respectivos custos de aluguel.

A análise inicial dos dados revelou uma variedade de fatores que podem influenciar os valores do aluguel de imóveis. Por meio de visualizações gráficas, conseguimos identificar possíveis padrões e correlações, fornecendo uma base sólida para nossa investigação, onde foi possível se perceber a veracidade dos dados e da capacidade do modelo.

A preparação dos dados desempenhou um papel crucial, incluindo a limpeza, transformação e normalização das variáveis. Ao aplicar técnicas como transformações logarítmicas para normalizar os dados e filtrar observações discrepantes, conseguimos melhorar a qualidade e coesão dos dados, preparando-os para a análise posterior.

Ao construir e avaliar modelos de aprendizado de máquina, exploramos diferentes abordagens para a previsão de preços de aluguel de imóveis. Focamos em identificar relações entre as características dos imóveis e os custos, permitindo que nossos modelos capturassem as nuances inerentes aos dados. Ao comparar os resultados entre diferentes modelos, pudemos avaliar a eficácia e a adaptabilidade de cada abordagem.

Este estudo reforça a ideia de que o aprendizado de máquina desempenha um papel crucial na previsão de preços de aluguel de imóveis, fornecendo *insights* valiosos para tomadas de decisões informadas por inquilinos do mercado imobiliário. Além disso, ressalta a importância da análise exploratória, pré-processamento de dados adequado e avaliação rigorosa de modelos para alcançar previsões mais precisas e confiáveis. No contexto de um mercado imobiliário em constante evolução, a aplicação de aprendizado de máquina oferece uma abordagem proativa para entender e antecipar as tendências de preços, auxiliando os inquilinos presentes neste mercado a tomar decisões embasadas e estratégicas.

## Referências

ALMEIDA, Cristian Machado de. Machine Learning: Conheça um pouco mais sobre o "aprendizado de máquina" e algumas de suas aplicações. 2020. Disponível em: <https://www.revistaferramental.com.br/artigo/machine-learning-conhecicapouco-maissobre-aprendizado-de-maquina-e-algumassuas-aplicacoes/>. Acesso em: 20 mar. 2023.

BARROS, Vinicius Gomes. O que devo analisar para comprar um imóvel com segurança? 2021. Disponível em: <https://www.jusbrasil.com.br/artigos/o-que-devo-analisar-para-comprar-um-imovel-com-seguranca/1186051953>. Acesso em: 6 mar. 2023.

COSTA, Marcia. O Mercado Imobiliário e sua relação com a economia. [s.d]. Disponível em: <https://mkavaliacoesimobiliarias.com.br/mercado-imobiliario-relacao-com-a-economia/>. Acesso em: 28 mar. 2023.

DIDÁTICA TECH. O que é Machine Learning (Aprendizado de Máquina). [s.d.]. Disponível em: <https://didatica.tech/o-que-e-machine-learning-aprendizado-de-maquina/>. Acesso em: 20 mar. 2023.

FACELI, Katti *et al.* Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. 2. ed. Rio de Janeiro: LTC, 2021.

FERNANDES, Antonio Carlos. Cuidados ao alugar um imóvel. 2019. Disponível em: <https://www.jusbrasil.com.br/artigos/cuidados-ao-alugar-um-imovel/795126847>. Acesso em: 12 mar. 2023.

KAGGLE. `brazilian_houses_to_rent`. 2020. Disponível em: <https://www.kaggle.com/datasets/rubenssjr/brazilian-houses-to-rent?resource=download>. Acesso em: 30 mar. 2023.

MATPLOTLIB. Matplotlib 3.8.0 documentation. 2012. Disponível em: <https://matplotlib.org/stable/index.html>. Acesso em: 22 mar. 2023.

MIRANDA, Vinicius Gouveia de. Validação de modelos lineares: Uma análise residual. 2016. Disponível em: [https://downloads.o-pitblast.com/files/technical\\_pt/Validacao%20de%20modelos%20-%20Uma%20analise%20residual.pdf](https://downloads.o-pitblast.com/files/technical_pt/Validacao%20de%20modelos%20-%20Uma%20analise%20residual.pdf). Acesso em: 22 mar. 2023.

MÜLLER, Andreas C; GUIDO, Sarah. Introduction to Machine Learning with Python: A Guide for Data Scientists. 1. ed. Sebastopol: O'Reilly Media, 2016.

NEOWAY. Machine Learning: Conceitos, definição e mais – Veja. 2020. Disponível em: <https://blog.neoway.com.br/inteligencia-artificial/#:~:text=Machine%20Learning&text=Como%20o%20pr%C3%B3prio%20nome%20j%C3%A1,chegar%20a%20uma%20determinada%20conclus%C3%A3o>. Acesso em: 10 mar. 2023.

PANDAS. Pandas - Python Data Analysis Library. [s.d]. Disponível em: <https://pandas.pydata.org/about/index.html>. Acesso em: 22 mar. 2023.

PURESTORAGE. O que é aprendizado de máquina (Machine Learning)? [s.d]. Disponível em: <https://www.purestorage.com/br/knowledge/what-is-deep-learning.html>. Acesso em: 10 mar. 2023.

PYTHON SOFTWARE FOUNDATION. [s.d]. Disponível em: <https://www.python.org/psf-landing/>. Acesso em: 9 set. 2023.

SAS. Machine Learning: O que é e qual sua importância? [s.d]. Disponível em: [https://www.sas.com/pt\\_br/insights/analytics/machine-learning.html](https://www.sas.com/pt_br/insights/analytics/machine-learning.html). Acesso em: 10 mar. 2023.

SCIKIT-LEARN. User guide: contents — scikit-learn 1.3.0 documentation. [s.d]. Disponível em: [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html). Acesso em: 22 mar. 2023.

SEABORN. User guide and tutorial — seaborn 0.12.2 documentation. [s.d]. Disponível em: <https://seaborn.pydata.org/tutorial.html>. Acesso em: 22 mar. 2023.