

MINICURSO – INTRODUÇÃO À RASPAGEM DE DADOS E INTELIGÊNCIA ARTIFICIAL NA ENGENHARIA DE AVALIAÇÕES

Eng. Bruno Henrique Gazzinelli

0. Programação do Curso



- **Momento 01** – **Introdução e Conceitos** de Raspagem de Dados;
- **Momento 02** – **Aplicações** de Raspagem e **Tratamento** de Dados
- **Momento 03** – **Estudo** de Caso e **Conclusões**

Momento 01 – Introdução e Conceitos de Raspagem de Dados

1. Introdução e Conceitos de Raspagem de Dados



Demonstração Técnica

1. Introdução e Conceitos de Raspagem de Dados



O que é? Web Scraping

1. Introdução e Conceitos de Raspagem de Dados

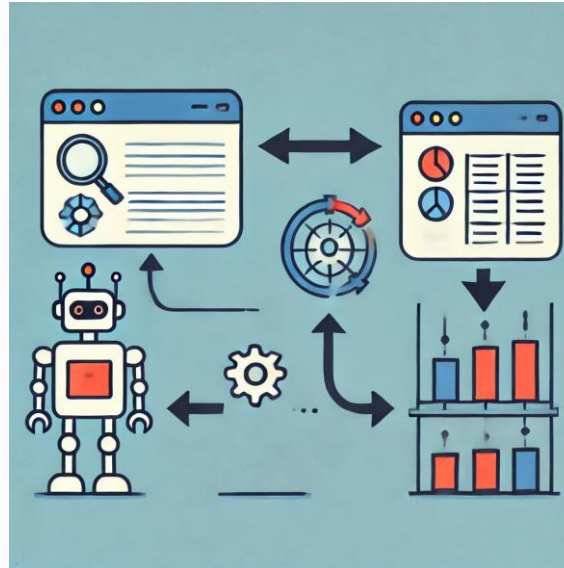


O que é Web Scraping ou Raspagem de Dados?

- Web scraping é o **processo automatizado** de coleta de dados de sites na internet.
- Ele permite a **extração de grandes volumes de informações** de páginas da web de forma eficiente, economizando tempo e esforço manual.
- Esses dados, frequentemente não estruturados (como HTML), são organizados em **formatos estruturados**, como planilhas ou bancos de dados, facilitando análises e estudos.
- O Web scraping é amplamente utilizado em setores como e-commerce, marketing, e avaliações imobiliárias para **analisar tendências de mercado e preços**.

1. Introdução e Conceitos de Raspagem de Dados

O que é Web Scraping ou Raspagem de Dados?



Fonte: Inteligência Artificial Generativa

1. Introdução e Conceitos de Raspagem de Dados

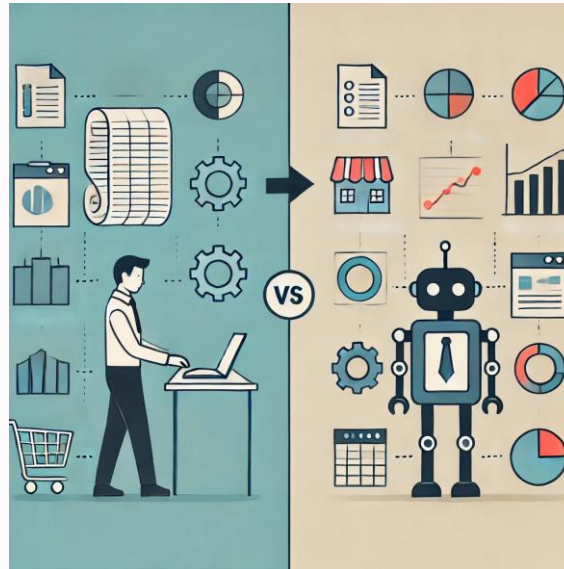


O que é Web Scraping ou Raspagem de Dados?

- Web scraping é amplamente utilizado em várias indústrias:
 - **E-commerce:** coleta de preços de produtos de diversos sites para comparação em tempo real.
 - **Imobiliário:** extração de dados de imóveis (preço, localização, características) para análise de tendências.
 - **Marketing:** monitoramento de menções de marcas em redes sociais e blogs.
- Ele permite capturar **grandes volumes de dados** de forma **rápida e eficiente**.

1. Introdução e Conceitos de Raspagem de Dados

O que é Web Scraping ou Raspagem de Dados?



Fonte: Inteligência Artificial Generativa

1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: Definição Técnica

- Web scraping é a técnica de **extração automatizada** de dados de sites, acessando páginas da web e "lendo" a **estrutura HTML**.
- A ferramenta identifica as **seções relevantes** (como textos, imagens, links) e organiza esses dados em **formatos estruturados** (CSV, JSON, etc.).
- O Web Scraping pode ser aplicado tanto em **páginas estáticas** quanto **páginas dinâmicas**, requerendo diferentes abordagens:

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Definição Técnica

Estruturação de Dados

- **Estruturados:** Representados de forma organizada em linhas e colunas, como em uma planilha ou banco de dados. São de fácil interpretação e processamento.
- **Semi-Estruturados:** Têm alguma organização, mas não seguem um formato rígido. Um exemplo seria um arquivo JSON ou XML.
- **Não Estruturados:** São desorganizados, como textos, vídeos, ou fotos, sendo mais difíceis de processar e organizar sem ferramentas adequadas



Fonte: images.google.com

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Definição Técnica

Página Estática

- Um site de **portfólio pessoal** ou **uma página institucional** de uma empresa onde o conteúdo é sempre o mesmo para todos os usuários e não muda com base em interações.
- O conteúdo HTML é **carregado uma vez e não muda** até que seja atualizado manualmente.



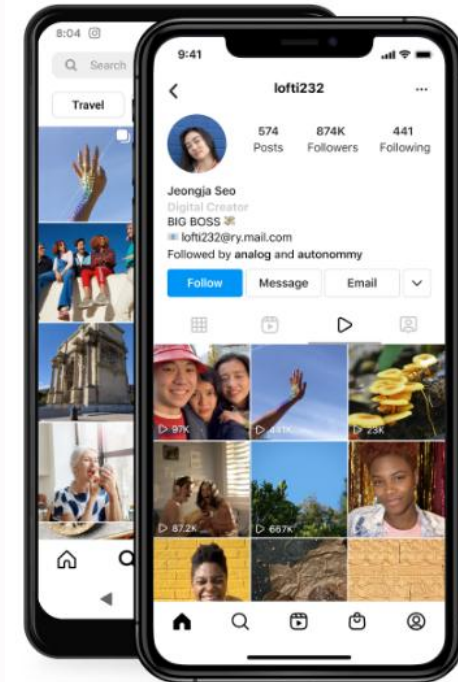
Fonte: www.ibape-nacional.com.br/site/

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Definição Técnica

Página Dinâmica

- Um site de **e-commerce** onde os preços e os produtos disponíveis mudam conforme a navegação, ou uma **rede social** onde novos posts são carregados à medida que o usuário rola a página.
- O conteúdo **muda dinamicamente** com base nas interações dos usuários ou dados de servidores.



Fonte: www.instagram.com

1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: Entendendo a estrutura HTML de uma página

- As páginas da web são estruturadas em **HTML** (HyperText Markup Language).
- O HTML organiza o conteúdo da página em **blocos**, como texto, imagens, e links.
- Web scraping "lê" essa estrutura e identifica as partes que contêm as informações desejadas.
- **Exemplo de tag HTML:** <table> organiza dados em forma de tabela, muito comum para exibir preços em sites.

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Entendendo a estrutura HTML de uma página

The screenshot displays a real estate listing for an apartment in Belo Horizonte. The page layout includes a header with navigation links, a main content area with property details, and a sidebar with a contact form. The browser's developer tools are open on the right, showing the HTML structure of the page.

Destaque

Venda / MG / Apartamentos à venda em Belo Horizonte /
Palmeiras / Rua Senhora do Porto

Condomínio
não informado

Venda
R\$ 450.000

Preço abaixo do mercado

75 m² 3 quartos 2 banheiros
2 vagas 12 andar 1 suite

[Todas as características](#)

Endereço
Rua Senhora do Porto, 1330 - Palmeiras, Belo Horizonte - MG

[Explore a localização do imóvel](#)

Envie uma mensagem

Nome

E-mail

Telefone

Olá, gostaria de ter mais informações para comprar:
Apartamento, R\$ 450.000, Rua
Senhora do Porto, 1330 - Palmeiras

☐ Gostaria de receber o contato de
anunciantes com imóveis similares.

Enviar mensagem

Ao "Ver telefone", "Enviar mensagem" e/ou "WhatsApp"
você aceita os [Termos de uso](#) e [Política de](#)

```
<script defer src="https://static.cloudflareinsights.com/beacon.min.js/vcd15cbe=" integrity="sha512-Zps0mlRQV6y907TI0dKBHq9Md29nnaEIP1kf84rnaERnq6zVvVPUqr2ft8MlaS28oN72PdrCz5jY4U6VaAw1EQ=" data-cf-beacon="{\"rayId\":\"8c76f5675869ae87\", \"serverTiming\":{\"name\":\"cfExtPri\":true, \"cfL4\":true}}, \"version\":\"2024.8.0\", \"token\":\"7bd6b93a4d984df8aebc5cfd242ac8ac\"}" crossorigin=\"anonymous\"></script>
<iframe src="https://quiz.grupoazag.com/" style="width: 0px; height: 0px; border: 0px; position: fixed; bottom: 100%;"></iframe>
<script type="application/ld+json"></script>
<script type="application/ld+json"></script>
<script type="application/ld+json"></script>
<header class="header header--transparent" data-testid="header-component"></header>
<div style="top: 56px; right: 20px; z-index: 9999; position: fixed;"></div>
<section aria-hidden="true" class="main-carousel-container" data-testid="carousel-container" data-cy="carousel-container"></section>
<div class="base-page__content-container"></div>
<div class="mobile-floating-buttons"></div>
<footer class="footer"></footer>
<script id="pbjs-fix" data-ns="afterInteractive">if (window) { window.pbjs = {} }</script>
<script src="https://www.googletagmanager.com/etags/js?id=UA-126375-31" async="true" data-ns="afterInteractive"></script>
<script id="clickstream-tracker" data-ns="afterInteractive"></script>
```

Fonte: <https://www.vivareal.com.br/imovel/apartamento-3-quartos-palmeiras-bairros-belo-horizonte-com-garagem-75m2-venda-RS450000-id-2743141780/>

1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: Como os dados são organizados?

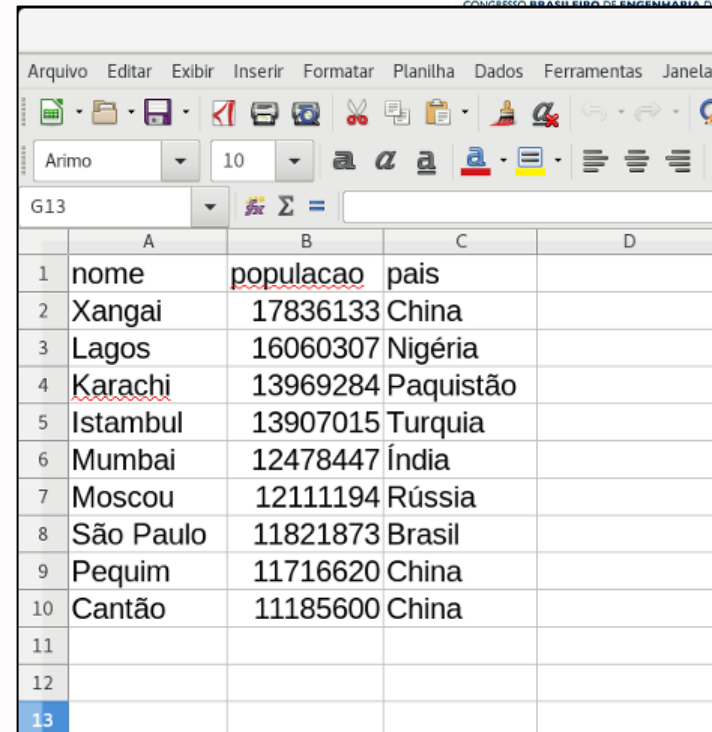
- Após a extração, os dados raspados precisam ser armazenados de forma **estruturada** para facilitar sua análise.
- Formatos comuns de armazenamento incluem:
 - ❖ **CSV:** Organiza os dados em linhas e colunas, como uma planilha.
 - ❖ **JSON:** Usado para armazenar dados complexos em uma estrutura hierárquica, útil para representar informações mais detalhadas.
- A escolha do formato depende da complexidade dos dados e do objetivo da análise.

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Como os dados são organizados?

CSV: Planilha de Dados

- Organiza os dados em linhas e colunas, como uma planilha.



The image shows a screenshot of a spreadsheet application. The menu bar includes Arquivo, Editar, Exibir, Inserir, Formatar, Planilha, Dados, Ferramentas, and Janela. The toolbar contains various icons for file operations and formatting. The active cell is G13. The data is organized in a table with 4 columns (A, B, C, D) and 13 rows (1 to 13). The first row contains headers: 'nome' in column A, 'populacao' in column B, and 'pais' in column C. The following rows contain data for various cities and countries.

	A	B	C	D
1	nome	populacao	pais	
2	Xangai	17836133	China	
3	Lagos	16060307	Nigéria	
4	Karachi	13969284	Paquistão	
5	Istambul	13907015	Turquia	
6	Mumbai	12478447	Índia	
7	Moscou	12111194	Rússia	
8	São Paulo	11821873	Brasil	
9	Pequim	11716620	China	
10	Cantão	11185600	China	
11				
12				
13				

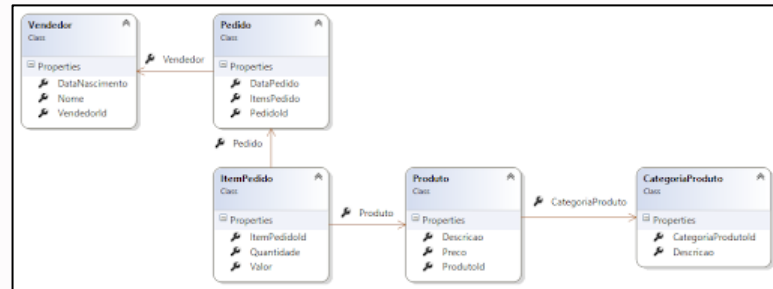
Fonte: www.images.google.com

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Como os dados são organizados?

JSON: Organização por “caixas”

- Usado para armazenar dados complexos em uma estrutura hierárquica, útil para representar informações mais detalhadas



Fonte: www.images.google.com

1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: Como o Web Scraping é realizado?

- O processo de web scraping envolve três etapas principais:
 - **Identificação de dados relevantes** na página (como preços, descrições, imagens).
 - **Análise do código HTML** da página para localizar essas informações.
 - **Extração automatizada** dos dados, que são organizados em um formato acessível (CSV, JSON, etc.).
- Ferramentas como **Selenium** e **BeautifulSoup** facilitam a navegação e a coleta de dados de páginas dinâmicas e estáticas.

1. Introdução e Conceitos de Raspagem de Dados



História e Evolução

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: História e Evolução



- Como surgiu o Web Scraping?

- O web scraping surgiu na **década de 1990**, conforme a internet crescia e havia a necessidade de coletar dados em grande escala.
- Inicialmente, as ferramentas eram **simples**, focadas em extrair dados de **páginas estáticas**.
- A popularização da Web 1.0 trouxe as primeiras soluções básicas de scraping, voltadas para **extração de texto e links**.

1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: História e Evolução

- Como a Web 2.0 mudou o Web Scraping?
 - A transição para a Web 2.0, com **páginas mais dinâmicas e interativas**, criou a necessidade de técnicas mais avançadas de scraping.
 - A evolução para sites com conteúdo dinâmico (redes sociais, e-commerce) exigiu **novas ferramentas** que pudessem lidar com a interação contínua de usuários.
 - Isso impulsionou a criação de frameworks como **Selenium**, que simula interações de usuários com páginas dinâmicas.

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: História e Evolução

Fonte: images.google.com



1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: História e Evolução

- Principais ferramentas ao longo do tempo
 - **BeautifulSoup**: Criado no início dos anos 2000, facilitou a extração de dados de HTML.
 - **Selenium**: Permitiu simular ações de usuários, crucial para scraping de páginas dinâmicas.
 - **Octoparse**: Democratizou o web scraping ao permitir que usuários sem habilidades de programação pudessem realizar raspagem de dados com uma interface visual.

1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: História e Evolução

- Restrições e Soluções Modernas

- Com o crescimento do web scraping, muitos sites implementaram políticas de restrição como o **robots.txt**, que instrui bots sobre as áreas da página que não podem ser acessadas.
- Alguns sites também utilizam **CAPTCHAs** ou limitam o número de acessos por endereço IP, dificultando a automação da coleta de dados.
- Ferramentas como **Selenium** permitem simular a navegação de um usuário real, enquanto o uso de **proxies rotativos** ajuda a contornar bloqueios baseados em IP.
- **Avanços recentes** incluem o uso de técnicas de **machine learning** para identificar e extrair dados de sites que alteram suas estruturas com frequência.

1. Introdução e Conceitos de Raspagem de Dados



Abordagem Legal e Ética

1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: Abordagem Legal e Ética

- Questões Éticas no Web Scraping

- Mesmo quando o scraping é tecnicamente permitido, ele pode levantar **questões éticas**.
- Coletar dados de um site sem permissão pode **prejudicar a reputação** de uma empresa ou **violar a confiança** dos usuários.

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Abordagem Legal e Ética



- Importância da Conformidade Legal

- O web scraping pode levantar questões legais e éticas, especialmente no que diz respeito à coleta e uso de dados.
- Leis de privacidade, como a **LGPD** (Brasil), **GDPR** (Europa), e **CFAA** (EUA), estabelecem diretrizes sobre a coleta, armazenamento e uso de dados pessoais.
- Empresas e profissionais devem estar atentos às regulamentações para evitar sanções legais e proteger a

1. Introdução e Conceitos de Raspagem de Dados



Web Scraping: Abordagem Legal e Ética

- O que dizem a IVSC e a ABNT?

- A IVSC e a ABNT **não se posicionam direta ou indiretamente** em relação ao Web Scraping ou qualquer outra forma automatizada de obtenção de dados.
- Entretanto, na IVS que entrará em vigor a partir de 31 de Janeiro de 2025, a IVSC faz menção as restrições ao ***Automated Valuation Model (AVM)***.
- “No model ***without the valuer applying professional judgement, for example an automated valuation model (AVM), can produce an IVS-compliant valuation.***”

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Abordagem Legal e Ética



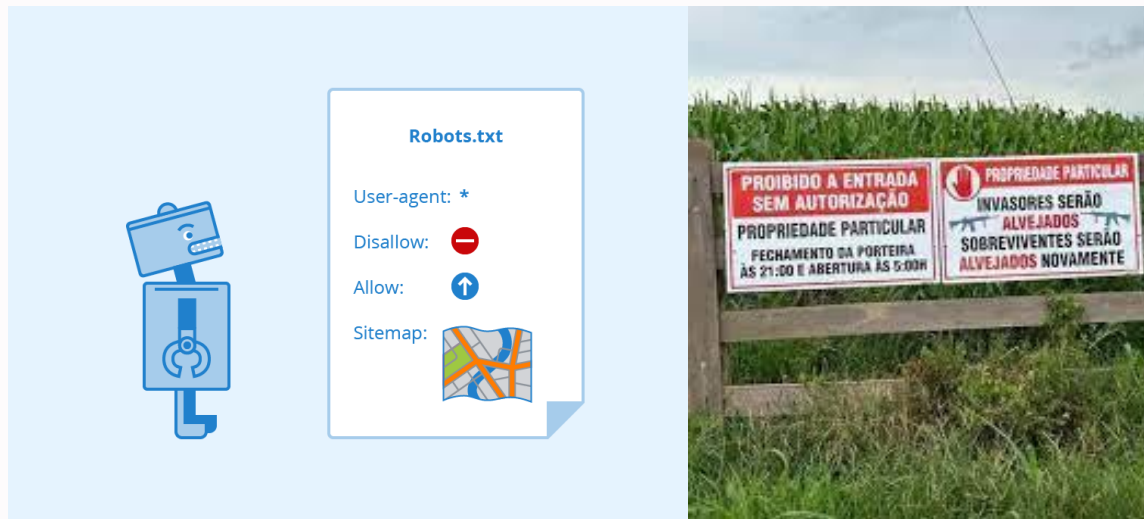
- Robots.txt e Limitações Legais

- O **robots.txt** é um arquivo usado por sites para instruir bots sobre quais áreas podem ser acessadas e quais não podem.
- Ignorar essas instruções pode ser visto como uma **violação de política do site**, resultando em sanções ou processos legais.
- O arquivo é uma prática comum para **proteger áreas privadas ou sensíveis** de um site.

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Abordagem Legal e Ética

➤ Diz o ditado: *“se tem placa, tem história”*



Fonte: images.google.com

1. Introdução e Conceitos de Raspagem de Dados

Web Scraping: Abordagem Legal e Ética



- Práticas recomendadas

- ✓ **Respeitar o arquivo robots.txt** para evitar a raspagem de áreas proibidas;
- ✓ **Solicitar permissão** para coletar dados sempre que possível;
- ✓ **Usar dados de maneira transparente** e com responsabilidade, garantindo a segurança da informação;
- ✓ **Usar proxies de maneira ética** para evitar sobrecarregar servidores de um único ponto de origem.

Momento 02 – Aplicações de Raspagem e Tratamento de Dados

2. Aplicações de Raspagem e Tratamentos de Dados



Técnicas e Ferramentas de Web Scraping

2. Aplicações de Raspagem e Tratamentos de Dados



Web Scraping: Técnicas e Ferramentas

Ferramentas Comerciais “*sem necessidade*” de Escrita de Códigos

- **Octoparse** – Octoparse Tech Co., Ltd;
- **WebScrapper** - Web Scraper Inc.;
- **Dataminer** – Skyline Communications;
- **Power Automate** – Microsoft Power Platform;
- **Instant Data Scraper** – Webrobots.io;

2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Técnicas e Ferramentas



1 - OctoParse

- O Octoparse é uma ferramenta de raspagem de dados que permite extrair informações de sites de **forma automática e sem necessidade de programação**;
- Na versão gratuita, oferece uma **interface intuitiva baseada em cliques**, suporte à **extração de listas e tabelas, paginação automática**, exportação para Excel ou CSV e execução local das tarefas;
- É ideal para projetos pequenos e para iniciantes em web scraping.
- Exemplo de aplicação: <https://www.youtube.com/watch?v=Zvysdas2bl0>

2. Aplicações de Raspagem e Tratamentos de Dados



Web Scraping: Técnicas e Ferramentas

2 - WebScraper

- Web Scraper é uma **extensão gratuita para os navegadores Chrome e Firefox**, usada para extrair dados estruturados da web;
- Permite configurar **sitemaps para navegar** por páginas automaticamente, coletando listas, textos, links e imagens;
- Os dados podem ser exportados em **CSV ou XLSX**;
- É ideal para tarefas recorrentes e de médio porte, com foco em usuários que preferem **soluções diretas no navegador**.
- Exemplo de aplicação: <https://www.youtube.com/watch?v=uVdNr-HT2ws>

2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Técnicas e Ferramentas



3 - DataMiner

- O DataMiner é uma extensão de navegador para **Chrome e Edge** que permite extrair dados de sites de forma automática, sem exigir conhecimentos de programação.
- Na versão gratuita, possibilita a **criação de receitas de extração** por meio de cliques nos elementos da página, além de oferecer **templates prontos criados pela comunidade**.
- Permite extrair **listas, tabelas, textos e links**, com exportação dos dados em formatos como Excel e CSV.
- Exemplo de aplicação: <https://www.youtube.com/watch?v=3zI7wTAK0KE>

2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Técnicas e Ferramentas



4 – Power Automate

- Power Automate é uma ferramenta da **Microsoft** voltada para **automação de processos**, incluindo tarefas de raspagem de dados simples.
- Por meio de **fluxos automatizados**, é possível configurar **bots que navegam por páginas**, coletam informações e as transferem para planilhas ou bancos de dados.
- Na versão gratuita (desktop), permite automatizar a navegação web com comandos visuais, **mas a raspagem de dados é limitada em comparação com ferramentas especializadas**.
- Exemplo de aplicação: <https://www.youtube.com/watch?v=OoMn0-YOJvY>

2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Técnicas e Ferramentas



5 – Instant Data Scraper

- Instant Data Scraper é uma **extensão gratuita** para navegadores baseada em automação de coleta de dados.
- A ferramenta **identifica automaticamente tabelas e listas** em páginas web, permitindo exportar as informações para CSV ou Excel.
- Possui recursos de **paginação e rolagem automática**, facilitando a captura de grandes volumes de dados em poucos cliques.
- Por ser uma solução rápida e sem necessidade de programação, é **indicada para coletas simples**, mas suas funcionalidades são limitadas em comparação com ferramentas de raspagem profissionais
- **Exemplo de aplicação:** <https://www.youtube.com/watch?v=iSzRyrDynBg&ab>

2. Aplicações de Raspagem e Tratamentos de Dados



Exemplos Práticos de Coleta de Dados

2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Conhecendo a Ferramenta



Octoparse: Ferramenta Visual de Web Scraping

- Octoparse é uma ferramenta de scraping que **não exige conhecimento de programação**, permitindo que qualquer pessoa configure tarefas de extração de dados de forma visual e intuitiva.
- Oferece uma **interface gráfica** para configurar fluxos de trabalho, tornando o **processo de scraping simples**, mesmo para páginas dinâmicas.
- Suporta uma **ampla variedade de sites**, extraindo dados de forma eficiente sem a necessidade de codificação.

2. Aplicações de Raspagem e Tratamentos de Dados



- **Web Scraping: Conhecendo a Ferramenta**
- **Octoparse:** Nossa escolha pelo **Octoparse** (www.octoparse.com), justificada pelas principais características listadas a seguir:
 - ✓ *Ausência de necessidade de escrita de código;*
 - ✓ *Facilidade de reconhecimento de elementos das páginas;*
 - ✓ *Possibilidade de agendamento de rotinas;*
 - ✓ *Possibilidade de progressão e reiteração entre links e páginas;*
 - ✓ *Alto grau de Personalização do fluxo de extração de dados.*

2. Aplicações de Raspagem e Tratamentos de Dados

Fonte: www.octoparse.com

The screenshot shows the Octoparse website interface. On the left is a dark blue sidebar with icons for Home, New, Dashboard, Template, Data Service, Tools, Inbox, Help, and Settings. The main content area has a light gray header with 'Home' and a greeting 'Hello, Octoparse. Glad to see you here!'. Below the greeting is a search bar with the placeholder text 'Search for a task template or input website URL(s)' and a 'Start' button. There are two main cards: 'What's a custom task?' with an 'otd' icon and 'What's a task template?' with a Google Assistant icon. Both cards have buttons for '+ New Task', 'Tutorial', and 'Templates'. To the right of these cards are two more cards: 'Easy Web Scraping for Anyone' with a play button and 'Basic Walkthrough' with a 'Let's go!' button. At the bottom, there's a 'Popular Templates' section with categories like Products, Social Media, Real Estates, Jobs, and Maps. It lists several templates: 'Product listing_Amazon' (Products | 620), 'Store reviews_Google...' (Maps | 582), 'Video list_YouTube' (Social Media | 466), 'Posts_ReddIt' (Social Media | 546), 'Search results_Yahoo!' (Search Engine | 478), and 'Search results_Google' (Search Engine | 518). On the right side of the page, there's a 'Resources' section with links for FAQ, Video, and Case, and a 'Tutorial' section with links for 'Octoparse 101' and 'XPath Beginners' Guide'.

2. Aplicações de Raspagem e Tratamentos de Dados

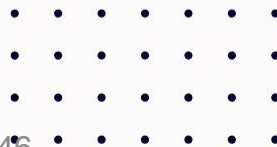
Fonte: www.octoparse.com

The screenshot displays the Octoparse website in a browser. The browser's address bar shows the URL: `octoparse.com/?utm_source=adsgoogle&utm_medium=search&utm_campaign=octoparse&gclid=CjwKCAjwu4Wo8h8kEwAojNdXp3g8EnBbDq/u...`. The website features a blue header with the Octoparse logo and navigation links: Solutions, Download, Pricing, and Resources. On the right side of the header are links for Login and a Start a free trial button. Below the header, there is a large section with the headline "No code is the best code". To the left of this text is a visual representation of the Octoparse web scraper interface, which includes a "Browse" tab, a "Setting" tab, and a "Save" tab. The "Browse" tab shows a preview of a website with product listings, including one for "Proton" with a price of \$49.90 and another for "Proton" with a price of \$89.99. The "Setting" tab shows a configuration window for selecting elements to scrape. The "Save" tab shows a "Go to Webpage" button. To the right of the headline, there is a paragraph of text: "Octoparse allows everyone to build reliable web scrapers they need - no coding needed. Design your own scraper in a workflow designer and get everything visualized in a browser." Below this text is a "Download now" button with a right-pointing arrow. The website also features a "Data Preview" section at the bottom left, which shows a table of data with columns for "Total" and "line of data".

2. Aplicações de Raspagem e Tratamentos de Dados



(01) - Bem vindo Octoparse!



2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Octoparse – Instalação

The screenshot displays the Octoparse website. The top navigation bar includes links for Solutions, Templates, Download, Pricing, and Resources, along with a 'Log in' button and a 'Start a free trial' button. The main content area is divided into two sections. On the left, the 'Download for Windows' section features a large heading, a download button for version 8.8.0, and a link to download for Mac. On the right, the 'Home' dashboard is visible, showing a greeting, a search bar for task templates, and several featured templates like 'Product listing_Amazon', 'Store reviews_Google...', and 'Video list_YouTube'. A cookie consent banner is at the bottom.

Octoparse Solutions Templates Download Pricing Resources Log in Start a free trial EN

Download for Windows

8.8.0

for Windows 7 (64-bit) or higher

Download for Mac

Home

Hello, Octoparse. Glad to see you here!

Search for a task template or input website URL(s) → Start

What's a custom task? Auto-detect webpages and set up custom workflows to extract data from any website. + New Task Tutorial

What's a task template? Choose from a selection of task templates and get data instantly with zero setup. Templates Tutorial

Easy Web Scraping for Anyone Learn Octoparse within minutes

Basic Walkthrough Get oriented with a step-by-step demo. Let's go! Task List

Popular Templates Products Social Media Real Estates Jobs Maps See All >

Product listing_Amazon Products 620

Store reviews_Google... Maps 582

Video list_YouTube Social Media 406

Posts_Reddit Social Media 14,946

Search results_Yahoo! Search Engines 4,236

Search results_Google Search Engines 516

Resources FAQ Video Case

[Tutorial] Octoparse 101

[Tutorial] XPath Beginners' Guide

This website uses cookies to ensure you get the best experience. Accept Close

2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Octoparse – Conhecendo a Interface

The screenshot displays the Octoparse web interface. On the left is a vertical sidebar menu with icons for 'Upg...', 'New', 'Task List', 'Templates', 'Tools', 'Pricing', 'RPA', 'Referral', 'Inbox', 'Help', and 'Settings'. The main content area is titled 'Hello, fvvbe73u. Glad to see you here!' and features a search bar for task templates or webpage URLs. Below the search bar are two cards: 'What's a custom task?' and 'What's a task template?'. The 'Popular Templates' section lists various scraping tasks with their respective icons and counts, such as 'Google Maps Email Finder' (124), 'Google Maps Scraper' (14), 'Google Maps Leads Scraper' (181), 'Google Maps Listings Scrape...' (64), 'Google Search Scraper' (66), 'Google Local Services Finder' (108), 'Contact Details Scraper' (62), 'Email & Social Media Scraper' (105), and 'LinkedIn Job Board Details S...' (35). On the right side, there is an 'Onboarding Checklist' with a progress bar and a 'Resources' section featuring a video titled 'How to Scrape IMDb Fast and Easy'.

Menu Lateral

Home | IRAP-SP | Apartamento...

Hello, fvvbe73u. Glad to see you here!

Tela Inicial – Home Page

Search for task templates or type webpage URL(s) → Start

What's a custom task?
Auto-detect webpages and set up custom workflows to extract data from any website.
+ New Task | Tutorial

What's a task template?
Choose from a selection of prebuilt task templates to get data instantly with zero setup.
88 Templates | Tutorial

Popular Templates | E-Commerce | Lead Generation | Social Media | Real Estate | See All >

- Google Maps Email Finder | Maps | 124
- Google Maps Scraper | Maps | 14
- Google Maps Leads Scraper ... | Maps | 181
- Google Maps Listings Scrape... | Maps | 64
- Google Search Scraper | Search Engine | 66
- Google Local Services Finder | Maps | 108
- Contact Details Scraper | Social Media | 62
- Email & Social Media Scraper | Social Media | 105
- LinkedIn Job Board Details S... | Jobs | 35

Onboarding Checklist **HOT**

- ☐ Start a Cloud Run | Earn \$2
- ☐ Schedule a Task | Earn \$2
- ☒ Complete Basic Walkthrough
- ☒ Run a Custom Task
- ☒ Complete a Data Export

Resources

How to Scrape IMDb Fast and Easy
From IMDb Pages to Excel

2. Aplicações de Raspagem e Tratamentos de Dados



Web Scraping: Octoparse – Conhecendo a Interface

Tela Inicial - Homepage

Ao iniciar o **Octoparse**, a tela inicial oferece uma visão geral das suas tarefas de Raspagem de Dados.

- **Navegar por Tarefas Pré Definidas:** Permite buscar e ter acesso rápido a templates pré-configurados para raspagem de dados.
- **Popular Templates:** Modelos populares para scraping de diferentes fontes (Google Maps, E-commerce, Social Media, etc.).
- **Tutoriais:** Acesso direto a tutoriais explicativos.

2. Aplicações de Raspagem e Tratamentos de Dados



Web Scraping: Octoparse – Conhecendo a Interface

Menu Lateral – Porção superior

Na lateral esquerda da Página Inicial do **Octoparse**, tem-se um menu de opções para navegação dentre as facilidades disponibilizadas pela ferramenta.

- **Criar Nova Tarefa:** Botão para iniciar um novo fluxo de trabalho.
- **Lista de Tarefas:** Exibe as últimas tarefas realizadas e seu status (completa, em execução, falha).
- **Templates:** Mostra notificações de tarefas e logs de execução.
- **Ferramentas:** Acesso a configurações adicionais para otimizar a execução de tarefas.

2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Octoparse – Criação de Nova Tarefa

Task Group: My Group New Group

Opções de Inserção de Links de Entrada

Opções de Criação de Tarefas

Custom Task
Template Task
Import Tasks
New Task Group

Enter manually | Import from file | Import from task | Batch generate

Enter or import one or more URLs here (line separated)

Please enter no more than 10K URLs.

Save

2. Aplicações de Raspagem e Tratamentos de Dados



Web Scraping: Octoparse – Criação de Nova Tarefa

Configurando uma Nova Tarefa no Octoparse

Ainda no Menu Lateral, tem-se a opção de criação de uma **nova tarefa/rotina**. Esta criação se dá a partir da escolha do ponto de partida da tarefa.

- **Custom Task:** Criar uma tarefa personalizada para raspagem de dados.
- **Template Task:** Usar modelos prontos para iniciar uma tarefa rapidamente.
- **Import Tasks:** Importar tarefas de fontes externas.
- **New Task Group:** Criar um grupo de tarefas para organizar fluxos de trabalho.

2. Aplicações de Raspagem e Tratamentos de Dados



Web Scraping: Octoparse – Criação de Nova Tarefa

Configurando uma Nova Tarefa no Octoparse

Ao iniciar a criação de uma **nova tarefa/rotina**, tem-se um ambiente para definir a fonte, o site que será alvo de raspagem de dados.

- **Task Group:** Agrupa suas tarefas para organização.
- **URL Source:** Escolha entre inserir URLs manualmente, importar de um arquivo ou gerar em lote.
- **URL Input:** Insira uma ou várias URLs (limite de 10.000 URLs por tarefa).
- **Salvar:** Após configurar a URL, clique em "Save" para iniciar a tarefa.

2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Octoparse – Lista de Tarefas

All Tasks

Running Cloud Tasks

Completed Cloud Tasks

Stopped Cloud Tasks

Filters(2)

Q Search

Refresh

Reset

☐ Task Name	Status	Next Run	Automatic Exp	Action
☐ IBAPE-SP - Apartamentos à venda em Belo Ho... My Group 2 hours ago	Run Run Ready Completed 0 lines	Not Set Not Set	Not Set Not Set	
☐ Apartamentos à venda em São Paulo, SP - Viv... My Group 17 hours ago	Run Run Ready Stopped 8094 lines (272 duplic...	Not Set Not Set	Not Set Not Set	
☐ Portal de Edificações Prefeitura de Belo Horiz... My Group 2 months ago	Run Run Ready Stopped 0 lines	Not Set Not Set	Not Set Not Set	
☐ Portal de Edificações Prefeitura de Belo Horiz... My Group 3 months ago	Run Run Ready Completed 3 lines	Not Set Not Set	Not Set Not Set	
☐ Apartamentos à venda em São Paulo, SP - Viv... My Group 4 months ago	Run Run Ready All Data Stopped 36 lines All Data	Not Set Not Set	Not Set Not Set	
☐ Apartamentos à venda em São Paulo, SP - Viv... My Group 4 months ago	Run Run Ready Stopped 28 lines	Not Set Not Set	Not Set Not Set	

Total 8 items, displaying 1-8

< 1 > 20 / page

Go to Page

2. Aplicações de Raspagem e Tratamentos de Dados



Web Scraping: Octoparse – Lista de Tarefas

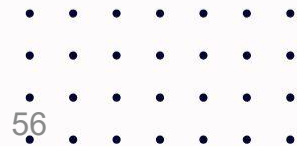
Tem-se a seguir um relatório das tarefas já executadas em momentos anteriores.

- **Task Name:** Nome da tarefa/rotina.
- **Status:** Apresenta se está em execução ou não, a quantidade de linhas criadas quando da execução e se há e quantas são as duplicatas.
- **Next Run:** Informação de Agendamento/Programação de repetição de execução daquela rotina.

2. Aplicações de Raspagem e Tratamentos de Dados



**Momento Interativo 01 –
Vamos observar!**



2. Aplicações de Raspagem e Tratamentos de Dados

Web Scraping: Octoparse – Ambiente de Trabalho

IBAPE-SP - Apartamento... Apartamentos à venda em Belo Horizonte, MG - Viva Real

EDITAR FILTROS

Dica

Visualização da Página

Rua Alexandre Siqueira, 60 - Salgado Filho, Belo Horizonte - MG
Apartamento com 2 Quartos à Venda, 47m²
47 m² 2 Quartos 1 Banheiro 1 Vaga
Aceita animais
R\$ 289.000
Condomínio: R\$ 220
TELEFONE ENVIAR MENSAGEM

Fluxo de Trabalho

Configurações Avançadas

Visualização Prévia dos Dados Extraídos

Data Preview Total [36] line(s) of data. Preview more data

Data Fields	No.	Title	Title_URL	Image	Image1	Image2	+	Actions
Page 1	1	Rua Alexandre Sique...	https://www.vivareal...	https://resizedimgs.v...	https://resizedimgs.v...	https://resizedimgs.v...	+	
Extract Data	2	Rua Alga Azul, 75 - J...	https://www.vivareal...	https://resizedimgs.v...	https://cdn1.vivareal...	https://resizedimgs.v...	+	
Extract Data1	3	Rua Leiloeiro José Ba...	https://www.vivareal...	https://resizedimgs.v...	https://cdn1.vivareal...	https://resizedimgs.v...	+	
	4	Rua Matipó, 290 - Sa...	https://www.vivareal...	https://cdn1.vivareal...	https://cdn1.vivareal...	https://cdn1.vivareal...	+	
	5	Rua Professor Costa ...	https://www.vivareal...	https://cdn1.vivareal...	https://cdn1.vivareal...	https://cdn1.vivareal...	+	

2. Aplicações de Raspagem e Tratamentos de Dados



Web Scraping: Octoparse – Como funciona o Fluxo de Trabalho

- **Visualização da Página:** Exibe o conteúdo do site em tempo real enquanto a tarefa é configurada.
- **Fluxo de Trabalho:** A árvore de comandos visualiza as etapas da raspagem.
- **Visualização Prévia dos Dados:** Mostra os dados extraídos antes da execução completa.
- **Configurações Avançadas:** Permite ajustes detalhados no processo de scraping.
- **Dicas:** Sugestões automáticas para otimizar o fluxo de trabalho.

Momento 03 – Estudo de Caso e Conclusões

00 – Apresentação



- Inicialmente, apresenta-se a seguir os **tópicos** que serão abordados neste breve momento:

1. Introdução e Objetivo do Estudo
2. Metodologia;
3. Estudo de Caso
4. Preparação da Modelagem;
5. Machine Learning;
6. Teste Prático – Random Forest;
7. Conclusões.

1 - Introdução e Objetivo do Estudo

Como e porque combinar métodos de *Web Scraping*,
Inteligência Artificial Descritiva e *Machine Learning*?

01 - Contextualização e Relevância do Estudo



- **Complexidade da Avaliação Imobiliária**

- A avaliação imobiliária é um processo que exige a análise detalhada de diversas variáveis para alcançar precisão e confiabilidade. Esse cenário destaca a importância da **coleta de dados** robusta e estruturada.

- **Aplicação do Web Scraping**

- Com o crescimento dos dados online, o Web Scraping torna-se uma ferramenta eficaz para a **coleta automatizada** de grandes volumes de dados, oferecendo informações detalhadas sobre o mercado imobiliário.

- **Integração de Inteligência Artificial**

- A inteligência artificial generativa, como o ChatGPT, possibilita a automação de tarefas, desde o **processamento de dados** até a geração de relatórios, tornando o processo de avaliação mais rápido e preciso.

01 - Objetivo do Estudo



- O objetivo principal deste trabalho é **demonstrar a viabilidade e a eficácia do uso de técnicas de Web Scraping** na coleta de dados para avaliações imobiliárias, bem como a aplicação de modelos de machine learning, especificamente o Random Forest, para prever o valor de imóveis.
- Além disso, busca-se **aliar a inteligência artificial generativa (ChatGPT)** como um ambiente que possibilite a execução de rotinas, processamento de dados e todas as outras tarefas envolvidas no processo de avaliação.

01 - Objetivo do Estudo

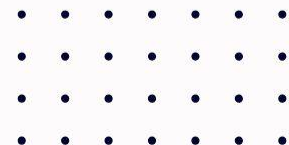


- Este estudo busca não apenas apresentar uma **metodologia prática e replicável** para avaliadores de imóveis, mas também fornecer insights sobre a importância da análise de dados e do uso de técnicas avançadas de machine learning e inteligência artificial generativa na avaliação imobiliária.

2 - Metodologia

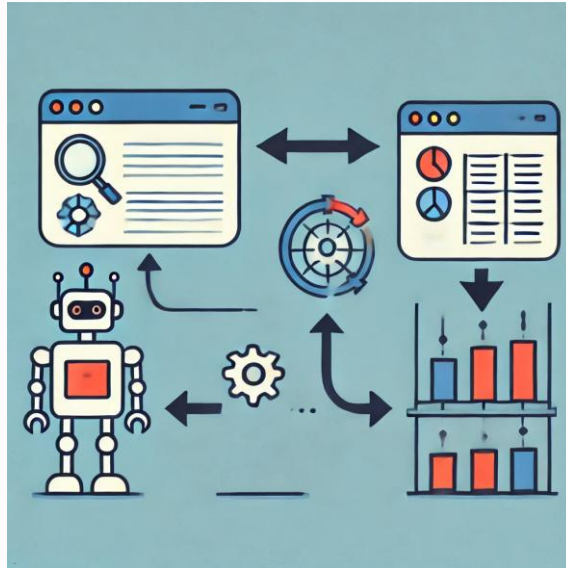
Detalha-se a seguir, a **metodologia desenvolvida** para o Estudo de Caso desenvolvido neste trabalho.

02 - Metodologia



02 – Escolha da Plataforma

- **Plataforma Utilizada: Octoparse**



- Para a Raspagem de Dados foi escolhida a plataforma “**Octoparse**”. Para tanto, foi adotada a versão “**v8.6.2.050811**”, na modalidade “**Free Plan**”.
- Trata-se de uma plataforma com solução pré-definida sem codificação para Web Scraping para transformar páginas em dados estruturados com cliques. Conta tanto com algoritmos pré-definidos de extração de dados, bem como da possibilidade de personalização para cada site, opção adotada para este trabalho.

02 - Fluxograma



3 – Estudo de Caso

Mercado Imobiliário Residencial de *imóveis de pequena metragem* em São Paulo – SP - Brasil

03 – Escolha de Tipologia



Definiu-se inicialmente a obtenção de dados de apartamentos que apresentassem áreas no intervalo de 20 a 50 m².

- I. **Grande disponibilidade** de dados;
- II. **Variabilidade** de plantas arquitetônicas;
- III. **Boa diferenciação de disposição** de dormitórios, banheiros e vagas;
- IV. **Grande variabilidade de comodidades** dispostas nos condomínios que estão instaladas e por fim;
- v. **Diversidade de localização** de logradouros e bairros em que se encontram.

03 – Refinamento e Tratamento

Classe
Textual

- Verificação de duplicidade de oferta através de verificação do campo de identificação da oferta.

Classe
Numérica

- Verificação de preenchimento incorreto ou inadequado dos campos numéricos;

Classe
Qualitativa

- Análise de declaração de campos e qualidade das informações.

03 – Amostragem Definitiva

Panorama Geral de Imóveis - Dormitórios	
	Dados da Amostragem Final
1 Dormitório	533,00
2 Dormitórios	507,00
3 Dormitórios	2,00
Total	1042,00

Panorama Geral de Imóveis - Banheiros	
	Dados da Amostragem Final
1 Banheiro	990,00
2 Banheiros	51,00
3 Banheiros	1,00
Total	1042,00

Panorama Geral de Imóveis - Vagas de Garagem	
	Dados da Amostragem Final
Não Especificado	424,00
1 Vaga	609,00
2 Vagas	9,00
Total	1042,00

- Amostra Final

- A partir das eliminações dos dados que possuíam inconsistências nas declarações, como resultado, obteve-se um total de 1042 (Mil e Quarenta e Dois).
- Assim, é importante pontuar que dos 1042 dados levantados, obteve-se um total de 256 bairros e 727 logradouros do município de São Paulo.

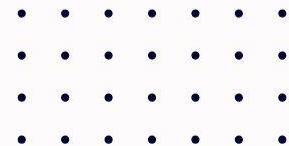
4 – Preparação da Modelagem

Realizada a coleta de dados, sua estruturação e tratamento, deu-se sequência à fase de **preparação para modelagem de dados, análise e processamento**, para posterior aplicação prática na avaliação de um imóvel.

04 – Integração com I.A. – L.L.M.



- Iniciou-se um processo de integração de processos a partir do uso do **Chat GPT 4o**, na versão paga do ambiente.
- Para tanto, estruturou-se um processo inicial de **alimentação de informações e diretrizes** para posterior prosseguimento das análises estatísticas.
- A integração fora efetivada na prática a partir de (i) **comandos diretos**, bem como do (ii) **upload de arquivos** em **.xlsx (Excel)** e **.pdf (Adobe Acrobat)** consolidando uma base de dados, bem como da definindo diretrizes técnicas.



04 – Fluxo de Análise

Fluxo de Análise de Dados de Regressão – Via ChatGPT 4o

Fluxo de Análise de Regressão conforme práticas de mercado & ABNT NBR-14653:2/2011

1. Início

Definição das Variáveis:

Verificar o comportamento isolado de cada variável independente em relação à variável dependente.

2. Análise das Variáveis

Identificação de Pontos Influentes.

Cálculo da Equação de Regressão.

Transformações BOX-COX: x , $\ln(x)$ ou $1/x$ (preferencialmente).

3. Verificação das Hipóteses

Analisar se as hipóteses formuladas estão ocorrendo.

4. Cálculo do Coeficiente de Determinação (R^2)

Representa o poder de explicação das variáveis independentes.

5. Cálculo do Coeficiente de Correlação (R)

Raiz quadrada de R^2 .

Indica a forma e a força da correlação existente entre as variáveis.

Correlação simples (2 variáveis) varia de -1 a 1.

6. Distribuição dos Resíduos

Verificar se os dados tendem ou não à curva normal para pequenas amostras.

Outliers: Menos de 5% dos dados.

7. Teste de Hipóteses

Verificar a significância do modelo.

8. Significância dos Regressores

Analisar a probabilidade de ocorrer erros ao se rejeitar uma hipótese que virá a ser verdadeira (probabilidade dos parâmetros a , b , c ,... serem ZERO).

9. Verificação de Homocedasticidade

Analisar se o comportamento da amostra está de acordo com a hipótese aferida pela Distribuição F de Snedecor.

10. Resíduos

Valor do dado – média calculada na equação para o dado.

Resíduos positivos indicam superavaliação e negativos indicam subavaliação.

Resíduo Relativo deve ser menor ou igual a 80% e desvio padrão menor ou igual a 2% (outliers).

Examinando dados com mais de 60% de resíduo: Reavaliar o dado.

Examinando dados com mais de 80% de resíduo: Preocupar-se com a aceitação do dado na amostra.

Mais de 100% de resíduo: Verificar o comportamento do modelo sem o dado.

11. Fim da Projeção dos Valores

Significância:

Grupo I: $\leq 10\%$

Grupo II: $\leq 5\%$

Grupo III: $\leq 1\%$

Análise de Sensibilidade: Verificar o comportamento da amostra conforme a hipótese.

04 – Regressão Linear Simples



- O objetivo principal da execução de regressões lineares simples foi **entender a relação entre cada variável independente e a variável dependente** (Valor Imóvel) de forma isolada.
- Isso nos permitiu identificar quais variáveis têm um **impacto significativo** no valor dos imóveis e quais podem ser **descartadas** ou necessitam de

transformação para melhor adequação ao modelo.

- A regressão linear seguiu, em ordem:
 - I. Preparação dos Dados;
 - II. Execução da Regressão Linear;
 - III. Avaliação dos Resultados;
 - IV. Análise da Importância das Variáveis Independentes



04 – Resultados – R.L.S.

Resultados da Regressão Linear Simples – Via Chat GPT 4o

Modelo

- Variável dependente: Valor do Imóvel

- Área: 6,735

- Variáveis independentes: Área, Dormitórios, Banheiros, Vagas

- Erro padrão: 991.897

- t: 6.791

- $P > |t|$: 0.000

Resultados

- Intervalo de confiança 95%: [4,789, 8,682]

- Constante (Intercepto): 396,700

- Erro padrão: 36,500

- Dormitórios: -255,000

- t: 10.878

- Erro padrão: 13,700

- $P > |t|$: 0.000

- t: -18.658

- Intervalo de confiança 95%: [325,000, 468,000]

- $P > |t|$: 0.000

- Intervalo de confiança 95%: [-282,000, -228,000]

04 – Resultados – R.L.S.

Resultados da Regressão Linear Simples – Via Chat GPT 4o

- Banheiros: 95,650

- Erro padrão: 25,500

- t: 3.751

- $P > |t|$: 0.000

- Intervalo de confiança 95%: [45,600, 146,000]

- Vagas: 56,410

- Erro padrão: 13,300

- t: 4.248

- $P > |t|$: 0.000

- Intervalo de confiança 95%: [30,300, 82,500]

Métricas do Modelo

- R-quadrado: 0.300

- R-quadrado ajustado: 0.297

- F-statistic: 111.2

- Prob (F-statistic): 7.04e-79

- Número de observações: 1,042

- Durbin-Watson: 1.793

Diagnósticos

- Omnibus: 671.664

- Prob(Omnibus): 0.000

- Jarque-Bera (JB): 12,316.410

- Prob(JB): 0.000

- Skew: 2.652

- Kurtosis: 18.986

04 – Interpretação dos Resultados



O modelo explica aproximadamente 30% da variação no valor do imóvel (R^2 -quadrado = 0.300).

A partir dos resultados apresentados, verificou-se que a Regressão Linear inicial, ora realizada pelo *ChatGPT*, não fora adequada para uso direto para determinação de um modelo estatístico sólido e que cumpra os preceitos técnico-normativos.

Mesmo tendo sido realizados ajustes de heterocedasticidade via teste de Breusch-Pagan, para o modelo log transformado, não foi possível prosseguir com a adoção de regressões lineares simples.

04 – Aprimoramento Técnico



- **Regressão Linear Múltipla**

- Incluir **múltiplas variáveis independentes** para capturar melhor as interações entre diferentes características dos imóveis. A regressão linear múltipla apresentou uma melhora no coeficiente de determinação (R^2).



- **Adoção de Técnicas de Regularização**

- Reduzir a complexidade do modelo e evitar overfitting. Ajudaram a **estabilizar os coeficientes de regressão**, reduzindo a influência de multicolinearidade e melhorando a robustez do modelo.

- **Análise de Resíduos e Transformações**

- Garantir que os resíduos do modelo **seguissem uma distribuição normal** e que a variância fosse constante (homocedasticidade).



04 – Aprimoramento Técnico



- **Testes de Significância dos Regressores**

- Verificar a **significância estatística** de cada variável independente no modelo. (Área, Dormitórios, Banheiros e Vagas de Estacionamento).



- **Validação Cruzada**

- Avaliar a **estabilidade** e a generalização do modelo.

- **Comparação de Modelos**

- Utilizamos métricas como R^2 ajustado, **AIC** (Akaike Information Criterion) e **BIC** (Bayesian Information Criterion) para avaliar a qualidade dos modelos.



5 – LLM - Machine Learning

Após a análise inicial com regressões lineares simples e múltiplas, e considerando os ajustes necessários para garantir a robustez dos modelos, avançamos para técnicas mais sofisticadas de Machine Learning. O objetivo era **melhorar a precisão das previsões do valor dos imóveis**, capturando as interações complexas entre as variáveis.

05 – Apresentação de Testes



Random Forest

- **Objetivo:** Utilizar um conjunto de árvores de decisão para capturar interações complexas e não lineares entre as variáveis.
- **Resultados:** Mostrou um aumento significativo na precisão das previsões. Múltiplas árvores de decisão para minimizar o erro e evitar overfitting.

Gradient Boosting

- **Objetivo:** Construir sequencialmente modelos de regressão para minimizar o erro residual do modelo anterior.
- **Resultados:** Apresentou uma alta precisão, mas foi mais propenso a overfitting comparado ao Random Forest.

05 – Apresentação de Testes



- **Support Vector Machine (SVM)**

- **Objetivo:** Encontrar um hiperplano que maximiza a margem de separação entre diferentes classes ou prever valores contínuos para regressão.
- **Resultados:** A precisão foi razoável, mas o tempo de treinamento foi mais longo e a interpretabilidade do modelo foi mais difícil.



- **K-Nearest Neighbors (KNN)**

- **Objetivo:** Utilizar a proximidade dos dados para prever os valores dos imóveis.
- **Resultados:** Simples de implementar, mas a precisão das previsões foi inferior a do Random Forest e Gradient Boosting.



05 – Modelo Mais Eficiente



Para determinar o modelo mais eficiente, comparamos os modelos utilizando várias métricas de desempenho.



- I. **Coeficiente de Determinação (R^2)** - Avaliamos a proporção da variabilidade do valor dos imóveis que cada modelo conseguiu explicar;
- II. **Erro Médio Absoluto (MAE)** - Medimos a média dos erros absolutos das previsões para avaliar a precisão do modelo;



05 – Modelo Mais Eficiente



III. **Erro Quadrático Médio (MSE)** - Cálculo a média dos quadrados dos erros para penalizar mais severamente grandes erros de previsão;



IV. **Raiz Quadrada do Erro Médio (RMSE)** - Forneceu uma métrica na mesma escala dos valores preditos, facilitando a interpretação;

V. **Validação Cruzada** - Demonstrou uma performance consistente através dos diferentes subconjuntos dos dados, minimizando o risco de overfitting.



05 – Modelo Mais Eficiente



- ✓ O Random Forest foi identificado como o modelo mais eficiente para prever o valor dos imóveis, devido à sua capacidade de **capturar interações complexas e não lineares entre as variáveis**, sua robustez contra overfitting e sua alta precisão nas previsões.
- ✓ Este modelo nos forneceu uma **base sólida para incorporar variáveis adicionais**, como **comodidades e bairro**, melhorando ainda mais a precisão das nossas avaliações imobiliárias.

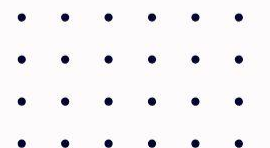
6 – Teste Prático – Random Forest

A adoção do modelo de Random Forest foi etapa crucial na análise, permitindo ***capturar interações complexas entre as variáveis*** e melhorar significativamente a precisão das previsões do valor dos imóveis.

06 – Teste Prático

- A amostra coletada continha **1042** registros de imóveis, com variáveis textuais, numéricas e qualitativas.
- **Variáveis Textuais:** Logradouro, Número, Bairro, Município, Estado, Chamada.
- **Variáveis Numéricas:** Área, Dormitórios, Banheiros, Vagas, Valor Imóvel, Condomínio.

- **Valor Previsto:** 51.346,54
- **Intervalo de Confiança (80%):**
 - **Limite Inferior:** U\$D 44.240,11
 - **Limite Superior:** U\$D 59.594,50
 - Percentual de **Amplitude do Intervalo** de Confiança: 29,90%



06 – Teste Prático – Análise Técnica



- O modelo **Random Forest** foi **eficaz** na previsão do valor dos imóveis, capturando interações complexas e não lineares entre as variáveis. Com um **R^2 de 0.82** no conjunto de teste,
- O modelo explicou uma **alta proporção da variabilidade** no valor dos imóveis, e os **erros de previsão (MAE e RMSE) foram razoavelmente baixos**.
- Este processo detalhado de avaliação utilizando Random Forest demonstrou ser robusto e preciso, proporcionando uma base confiável para a avaliação imobiliária e oferecendo insights valiosos para futuras análises e decisões de mercado.

7 – Conclusões

O uso de técnicas de Web Scraping aliado à inteligência artificial generativa e ao machine learning oferece uma **abordagem inovadora e eficaz para a avaliação imobiliária.**

07 – Riscos e Desafios

Qualidade dos Dados: A qualidade dos dados coletados via Web Scraping pode variar, dependendo da fonte e da estrutura dos dados nos portais de anúncios. Dados incompletos ou inconsistentes podem impactar negativamente a precisão das avaliações.

Manutenção e Atualização dos Algoritmos: A manutenção contínua dos algoritmos de Web Scraping é necessária para garantir que eles permaneçam eficazes diante de mudanças na estrutura dos sites de onde os dados são extraídos. Isso requer conhecimento técnico e recursos adicionais.

07 – Riscos e Desafios



Interpretação dos Resultados: A complexidade dos modelos de machine learning pode dificultar a interpretação dos resultados para avaliadores sem formação técnica avançada. A importância de variáveis pode ser obscurecida por interações complexas capturadas pelos modelos.



Conformidade com Normas e Regulamentações: Garantir que todos os procedimentos e técnicas estejam em conformidade com as normas técnicas e regulamentações vigentes, como a NBR 14.653 e a LGPD, é essencial para a validade e aceitação dos resultados das avaliações.



07 – Oportunidades



Automatização da Coleta de Dados: A utilização de Web Scraping permite a coleta automatizada de grandes volumes de dados de portais de anúncios imobiliários, garantindo uma amostra abrangente e atualizada. Isso reduz significativamente o tempo e os recursos necessários para a coleta manual de dados.

Melhoria na Precisão das Avaliações: A integração com técnicas de machine learning, como Random Forest, permite capturar interações complexas entre as variáveis, resultando em previsões mais precisas dos valores dos imóveis. A importância das variáveis foi claramente destacada, demonstrando a relevância das características físicas dos imóveis e das comodidades na determinação de seus valores.

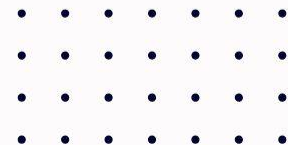
07 – Oportunidades

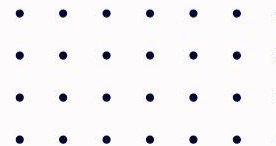


Análise Detalhada e Customizada: A inteligência artificial generativa, como o ChatGPT, oferece um ambiente que facilita a execução de rotinas de processamento de dados, análise estatística e geração de relatórios automatizados. Isso permite uma análise mais detalhada e customizada, adaptando-se às necessidades específicas de cada avaliação.



Eficiência e Redução de Custos: A automação dos processos de coleta e análise de dados reduz significativamente os custos operacionais e aumenta a eficiência, permitindo que os avaliadores se concentrem em aspectos mais estratégicos e complexos das avaliações imobiliárias.





BÔNUS – O pode ser o futuro?

BÔNUS – O que pode ser o futuro?

COBREAP - Quinta Feira – 20/11



📍 Auditório 01

10:30 - 12:00

MESA REDONDA

MR13 - INTELIGÊNCIA ARTIFICIAL NA ENGENHARIA DE AVALIAÇÕES: DESAFIOS E OPORTUNIDADES.



LUCIANO VENTURA (PR)

Moderador(a)



ANTÔNIO PELLI NETO

Debatedor(a)



BRUNO HENRIQUE SOARES GAZZINELLI CRUZ (MG)

Debatedor(a)

📍 Auditório 01

16:30 - 17:30

PALESTRA

P05 - RASPAGEM DE DADOS X PESQUISA HUMANA: ANTAGONISMO E COMPLEMENTAÇÃO



MARCOS MANSOUR CHEBIB AWAD (SP)

Apresentador(a)



BRUNO HENRIQUE SOARES GAZZINELLI CRUZ (MG)

Palestrante



AGRADECIMENTO

Bruno Henrique Gazzinelli



(31) 99300-9609



bruno@bhgengenharia.com

