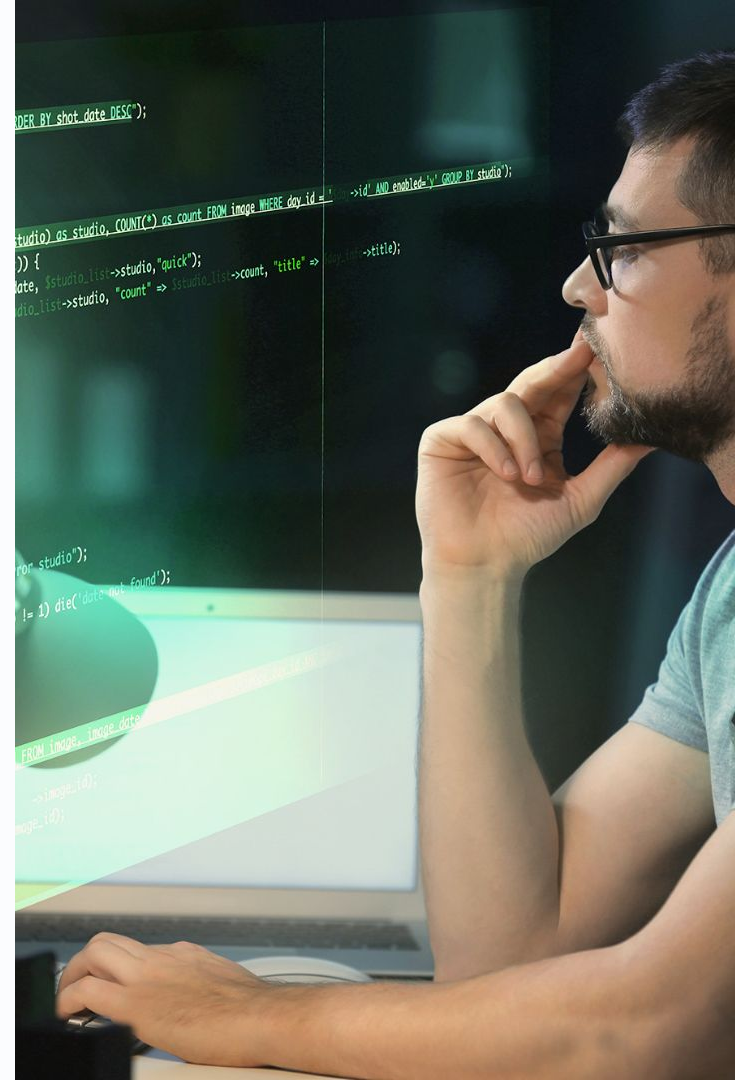


# Web Scraping:

Web scraping adalah sebuah cara pengambilan suatu data atau informasi tertentu dengan jumlah besar untuk nantinya digunakan dalam berbagai keperluan seperti riset, analisis dan lainnya

**Oleh Kharil Ilmi**

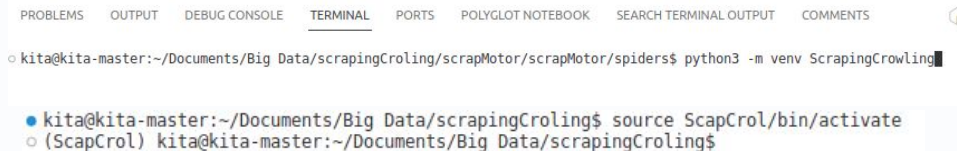


# Langkah Persiapan

## Langkah 1: Pasang dan aktifkan environment Virtual

“python3 -m venv ScapCrol”

“source env/bin/activate”



The screenshot shows a terminal window with a dark background. At the top, there are tabs for 'PROBLEMS', 'OUTPUT', 'DEBUG CONSOLE', 'TERMINAL' (which is active), 'PORTS', 'POLYGLOT NOTEBOOK', 'SEARCH TERMINAL OUTPUT', and 'COMMENTS'. The terminal output shows the following sequence of commands and their results:

```
kita@kita-master:~/Documents/Big Data/scrapingCroling/scrapMotor/scrapMotor/spiders$ python3 -m venv ScrapingCroling
kita@kita-master:~/Documents/Big Data/scrapingCroling$ source ScapCrol/bin/activate
(ScapCrol) kita@kita-master:~/Documents/Big Data/scrapingCroling$
```

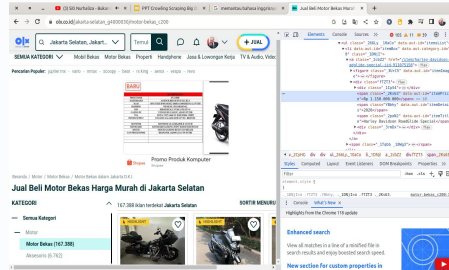
## Langkah 2: pasang scraping

'pip install Scrapy'.

# Langkah Analisa WEB



Langkah 1: navigasi ke situs web OLX Indonesia  
pilih kategori 'motor bekas'  
untuk mencari motor bekas.



Langkah 2: Inspect element yang dibutuhkan  
Inspect element untuk menjelajahi kode source halaman web.

```
<div class="fTzT3">
  <div class="_1IpS4">
    <span class="_2Ks63" data-aut-id="itemPrice">Rp 1.150.000.000</span> == $0
    <span class="YBbhy" data-aut-id="itemDetails">2020</span>
    <span class="_2poNJ" data-aut-id="itemTitle">Harley Davidson RoadGlide Special</span>
  <div class="_3rmDx">
```

Langkah 3: mengidentifikasi Data untuk mengikis  
Identifikasi data mana yang ingin discrap.

# Langkah coding spider

```
...
Robots.txt
User-agent: *
Disallow: /api/
Disallow: /post/
Disallow: /edit/
Disallow: /account/
Disallow: /chat/
Disallow: /profile/
Disallow: /payments/
Disallow: /nf/
Disallow: /items$
#Base Filters
Disallow: /*?*filter=*price_
Disallow: /*?*filter=*condition_
#Cars Filters
Disallow: /*?*filter=*cartype_
Disallow: /*?*filter=*mileage_
Disallow: /*?*filter=*transmission_
Disallow: /*?*filter=*year_

# Generated on 2023-07-25T10:40:38.725Z
...
```

## Langkah 1: Buat proyek kerokan baru

scrapy startproject projectname

## Langkah 2: Tentukan Data

Tentukan data yang diekstrak menggunakan XPath atau CSS.

## Langkah 3: Jalankan Spider dan simpan Data

'scrapy crawl motor -O motorJson1.csv.



```
1  import scrapy
2
3
4  class MotorSpider(scrapy.Spider):
5      name = 'motor'
6      start_urls = ['https://www.olx.co.id/jakarta-selatan_g4000030/motor-bekas_c200']
7
8      def parse(self, response):
9          for motor in response.css('div.fTZT3'):
10              try:
11                  yield{
12                      'harga':motor.css('span._2Ks63::text').getall(),
13                      'tahun':motor.css('span.YBbhy::text').getall(),
14                      'merk': motor.css('span._2poNJ::text').getall(),
15                  }
16              except:
17                  yield{
18                      'harga':'tidak ditemukan',
19                      'tahun':'tidak ditemukan',
20                      'merk': 'tidak ditemukan',
21                  }
22          nextBarang = response.xpath('//*[@id="main_content"]/div/div/section/div/div/div[5]/div[2]/div/div[2]/ul/li[22]/div/button')
23          if nextBarang is not None:
24              yield response.follow(nextBarang, callback=self.parse)
25
```

# Manfaat dari scraping Web

## 1 Pengambilan Keputusan

Membuat untuk membuat keputusan bisnis berdasarkan data yang akurat.

## 2 Keuntungan kompetitif

Bisa melacak aktivitas situs web perusahaan saingan.

## 3 Penghematan waktu dan biaya 🕒

Pengumpulan data dapat diotomatisasi, dan waktu dan biaya dapat lebih hemat.

# Aplikasi dunia nyata

1

## E-commerce

Scraping Web dapat membantu memantau harga pesaing dan melacak ketersediaan produk untuk menginformasikan harga dan keputusan inventori.

2

## Penelitian pasar

Kumpulkan data di preferensi konsumen dan tren dengan mengubah media sosial dan meninjau situs.

3

## Analisis keuangan

Lacak Harga stok dan gerakan pasar dengan menggores situs web berita keuangan.



Berejen