

Implementasi *Limit Python*

Teori, Studi Kasus Dan Implementasi Menggunakan *Python*

Mustofa Isnu Saherman¹⁾, Nur Budi Santhoso²⁾, Fathir Zulkarnain³⁾, Rizky Hermawan⁴⁾

^{1,2,3,4)}Universitas Pamulang, Tangerang Selatan

Jl. Surya Kencana No.1, Pamulang Bar., Kec. Pamulang, Kota Tangerang Selatan, Banten 15417

²⁾budi.rastafara@gmail.com

Abstract

This article discusses the application of the concept of Constraints in Python to control loops, variable values, and functions. We also explain the use of Constraints in machine learning, particularly in credit analysis, to control the amount of data used and improve model performance. Each topic is equipped with Python code examples to make it easier to understand and implement these concepts. By understanding and using Constraints, you can write code that is more efficient and easier to maintain, especially in managing literacy and variable values in programming and machine learning.

Keywords: Python, machine learning, literacy, constraints

1. Pendahuluan

Machine Learning (ML) telah meresap dalam kehidupan sehari-hari, membawa kenyamanan seperti perintah suara pada *smartphone* dan aplikasi diagnostik medis. Buku ini disusun praktis dengan contoh implementasi *Python*, cocok untuk pemula. Teori dasar disajikan singkat namun padat dengan ilustrasi yang mudah dipahami.

Buku ini terstruktur dalam empat kategori: Pengenalan ML, Teori Pendukung ML, ML *Lifecycle*, dan Algoritma-Algoritma ML. Penulis berterima kasih kepada semua yang mendukung pembuatan jurnal ini, terutama Universitas Pamulang. Kritik dan saran diterima dengan baik untuk peningkatan. Semoga jurnal ini memberikan manfaat dan inspirasi kepada pembaca, khususnya generasi yang ingin menjadi tanggap dan tangguh di era *Machine Learning*.

2. Metode Penelitian

2.1. *Maching learning*

ML merupakan bidang yang fokus pada desain dan analisis algoritma untuk memungkinkan komputer belajar tanpa kode spesifik. Algoritma generic dapat menghasilkan hal menarik dari data tanpa penulisan kode spesifik. Sebagai contoh, algoritma pengenalan tulisan tangan dapat digunakan untuk memfilter email spam tanpa perlu mengubah kode.

Definisi ML menurut Samuel (1959) adalah komputer belajar tanpa diprogram secara eksplisit. Mitchel (1997) menambahkan bahwa ML melibatkan *Task* (tugas), *Performance Measure* (ukuran kinerja), dan *Experience* (pengalaman).

Tugas ML melibatkan klasifikasi (pengelompokan data), regresi (memprediksi angka), transkripsi (mengubah data tidak terstruktur), *machine translation* (menerjemahkan

bahasa), *anomaly detection* (mendeteksi ketidaknormalan), dan *synthesis/sampling* (membuat sesuatu berdasarkan contoh).

Evaluasi performa ML berbeda sesuai tasknya, misalnya, klasifikasi dinilai berdasarkan tebakan benar/salah, sementara regresi dinilai berdasarkan kedekatannya dengan nilai asli.

Model ML memperoleh pengalaman dari dataset pelatihan, yaitu kumpulan contoh untuk dipelajari oleh komputer guna menyelesaikan tugasnya. Detail dataset akan dibahas lebih lanjut pada bab selanjutnya.

2.2. Pengumpulan data dengan metode *scrapping* berita detik.com

Sebagai contoh kita ingin mengambil data berita dari detik.com. Data yang diambil adalah data berita yang berisi judul, isi berita dan tanggal. Kita tidak mungkin mencatat satu per satu *item* yang ada di detik.com. Kita dapat membuat sebuah robot yang akan mengumpulkan data yang kita inginkan.

Agar mempermudah, pada kasus ini penulis akan menggunakan *Jupyter Notebook* sebagai tools untuk melakukan *scrapping*. Adapun langkah-langkah yang dilakukan adalah :

A. *Install library* yang dibutuhkan

Untuk melakukan *scrapping*, dibutuhkan beberapa *library* diantaranya *Beautiful Soup* dan *Request*. Untuk melakukan instalasi dapat menggunakan perintah berikut :

```
[1] 1 |pip install BeautifulSoup4  
    2 |pip install requests
```

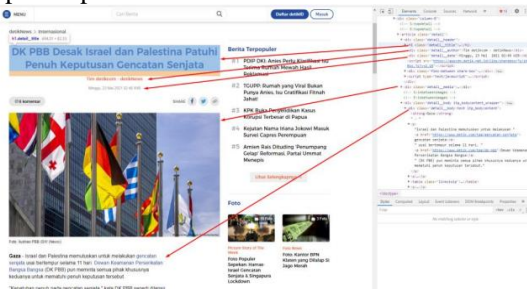
Gambar 1. Install Library yang Dibutuhkan

Setelah *library* terinstall semua maka kita sudah siap melakukan *scrapping*.

B. Mengekstrak isi Halaman *Web*

Inspeksi bertujuan untuk mengetahui elemen apa yang harus dipanggil untuk melakukan ekstraksi konten. Sebagai contoh satu halaman berita detik.com. Dengan bantuan *tool inspect* dari *google*

chrome maka kita dapat menganalisa struktur halaman berita detik.com. Untuk mengetahui harus melihat *source code* atau menggunakan *tool* inspektor pada Gambar.



Gambar 2. Mengekstrak isi Halaman Web

Hasil inspeksi menunjukkan bahwa :

- Judul berita terletak pada sebuah tag H1 dengan class *detail__title*
- Penulis terletak pada sebuah tag DIV dengan class *detail__author*
- Tanggal terletak pada sebuah tag DIV dengan class *detail__date*
- Isi berita terletak pada DIV dengan class *detail__body-text*, namun didalamnya ada iklan-iklan ke halaman lain dalam bentuk tabel, oleh karena itu iklan ini harus dibersihkan.

C. Parsing halaman web page menggunakan BeautifulSoup

Setelah mengetahui struktur html maka langkah selanjutnya adalah membuat kode yang memarsing halaman web tersebut. Setelah melakukan request pada URL tertentu (baris 9), maka selanjutnya adalah menganalisa hasil request menggunakan BeautifulSoup (baris 10). Untuk mencari Judul maka perlu menemukan element berdasarkan tag H1 dengan class *detail__title* (baris 12), tanggal dan author pada baris 13 dan 14 . Selanjutnya adalah pembersihan konten dilakukan dengan menghapus semua elemen tabel pada teks.

```

1 import requests
2 from bs4 import BeautifulSoup
3 import pandas as pd
4 import numpy as np
5
6 #ambil berita detik
7 def getBeritaDetik(url):
8     B = {}
9     response = requests.get(url)
10    soup = BeautifulSoup(response.text, 'html.parser')
11    #ambil elemen-elemen berita
12    B['judul'] = soup.find('h1', {'class': 'detail__title'}).text.replace('\n', '').strip()
13    B['tanggal'] = soup.find('div', {'class': 'detail__date'}).text.replace('\n', '').strip()
14    B['author'] = soup.find('div', {'class': 'detail__author'}).text.replace('\n', '').strip()
15    berita = soup.find('div', {'class': 'detail__body-text'})
16    text_berita = berita.text
17    #bersihkan isi berita
18    blah = berita.find_all("table")
19    for x in blah:
20        text_berita = text_berita.replace(x.text, '').replace('\n', '').strip()
21    #print(x.text)
22    B['berita'] = text_berita
23    return(B)

```

Gambar 3. Mengekstrak Isi Halaman Dibutuhkan

Method tersebut dapat digunakan untuk mengambil 1 halaman berita contohnya:

```

1 getBeritaDetik("https://news.detik.com/berita/"+
2 "d-5579822/detik-detik-mobil-porsche-putih-terbakar-di-kelapa-gading")

```

```

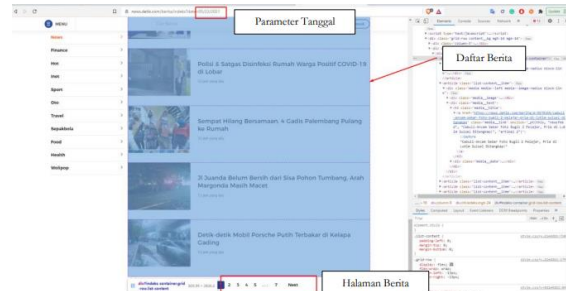
{'author': 'Yogi Ernes - detikNews',
 'berita': 'Jakarta - Sebuah mobil sport merek Porsche hangus terbakar di Kelapa Gading, Jakarta Utara.',
 'judul': 'Detik-detik Mobil Porsche Putih Terbakar di Kelapa Gading',
 'tanggal': 'Sabtu, 22 Mei 2021 19:49 WIB'}

```

Gambar 4. Contoh

D. Mengambil index daftar berita pertanggal

Setelah berhasil mengambil data per berita, maka langkah selanjutnya adalah mengambil seluruh berita. *Index* berita tersedia pada alamat <https://news.detik.com/berita/index> dan dapat dilihat berdasarkan tanggal. Daftar berita pada tanggal tertentu dipisahkan berdasarkan halaman-halaman, karena jumlah berita pada tanggal tersebut cukup banyak. Hasil analisa menunjukkan bahwa kita dapat memberikan parameter date dan nomer halaman. Struktur halaman *index* dapat dilihat pada Gambar 5.



Gambar 5. Halaman index

Semua konten terletak pada div dengan kelas *list-kontent* dan judul dan url terletak pada tag *article*. Setelah mengambil list url yang ada pada artikel, maka berdasarkan *link* tersebut kita ambil detail beritanya (baris 21). Adapun *source code* untuk mengambil list adalah sebagai berikut

```

1 import requests
2 from bs4 import BeautifulSoup
3
4 def indexBerita (tanggal, jumlahHalaman):
5     daftarBerita = []
6     halaman=0
7     for halaman in range(0, jumlahHalaman):
8         halaman = halaman + 1
9         base_url = "https://news.detik.com/berita/index/" + str(halaman)+"?date="+tanggal
10        #print(base_url)
11
12        # Request URL and BeautifulSoup Parser
13        r = requests.get(base_url)
14        soup = BeautifulSoup(r.text, "html.parser")
15
16        berita_container = soup.find('div', {'id': 'indeks-container'})
17        berita = berita_container.find_all('article')
18
19        for item in berita:
20            x = item.find("a", href=True)
21            berita = getBeritaDetik(x['href'])
22            daftarBerita.append(berita)
23        return daftarBerita
24

```

Gambar 6. Kode program

Luaran dari *method* tersebut dapat kita oleh lebih lanjut atau disimpan melalui *excel* menggunakan *pandas*

```

1 tanggal = '05/22/2021'
2 jumlahHalaman=1
3 list_berita = indexBerita(tanggal, jumlahHalaman)
4 Solah pada dataframe
5 df = pd.DataFrame(list_berita)
6 print(df.columns)
7 print(df)

```

```

Index(['judul', 'tanggal', 'author', 'berita'], dtype='object')
0  Menhub: Masyarakat dari Sumatera ke Jawa Wajib...  Jakarta - Menteri Perhubungan (Menhub) Budi Ka...
1  Cabuli-Ancam Sebar Foto Bugil 2 Pelajar, Pria...  Luwu Timur - Pemuda bernama Adrian (22) di Luw...
2  Mendes Minta Pendamping Desa Gotong-Royong Ban...  Jakarta - Menteri Desa, Pembangunan Daerah Ter...
3  Gempa M 3,6 Terjadi di Saingapu Sumba Timur...  Jakarta - Gempa berkekuatan magnitudo (M) 3,6...
4  Beri Benih hingga Cold Storage, Mentan Harap P...  Jakarta - Kementerian Pertanian (Kemantan) ser...
5  MUI Setuju Saran JK soal Kontak Amal untuk Pale...  Jakarta - Majelis Ulama Indonesia (MUI) mememin...
6  Dalih Ngentuk Pemobil Tabrak Lari Pedagang Mi...  Jakarta - Berkat "kesaktian" teknologi, penge...
7  Mendes Jadikan 2 Desa di Jatim Percontohan Pem...  Jakarta - Menteri Desa, Pembangunan Daerah Ter...
8  KPK Rantah OTT Bupati Nganjuk Ditangani Barek...  Jakarta - Direktur Sosialisasi dan Kampanye An...
9  Tejak Dokter Asal Sumut Jual Vaksin Corona 1la...  Jakarta - Polia Sumatera Utara (Sumut) meneta...
10  Cegah Lonjakan COVID-19, Satgas Minta 7 Kota Z...  Jakarta - Satgas Penanganan COVID-19 meminta p...
11  Libur Lebaran, Vaksinasi Lansia di NTB Tetap J...  Jakarta - Vaksinasi lansia tetap dilaksanakan ...
12  Saling Serang Partai Umat-Ngahalin Buntut Am...  Jakarta - Tenaga Ahli Utama Kantor Staf Presi...
13  Tak Diizinkan, Pulhan Pesepeda Galai Peringat...  Jakarta - Beberapa organisasi kemasyarakatan s...

```

Gambar 7. Hasil penyimpanan

E. Mengubah dataset menjadi Excel

Untuk mengubah menjadi Excel maka cukup menjalankan perintah `to_excel` pada `Pandas DataFrame` anda. Contoh penggunaannya adalah berikut

```
1 df.to_excel("berita.xls")
```

A	B	C	D	E	F
	judul	tanggal	author	berita	
0	Meriah! Masyarakat Sabtu, 22 Mei 2021 23:41	Matus Alfons - detikNews	Jakarta	Menteri Perhubungan (Menhub) Budi Karya Sumadi meminta ag	
1	Cabul-Ancam Sebar Sabtu, 22 Mei 2021 23:11	Herawan Mappawali - cLuwu Timur		Pemuda bernama Adnan (22) di Luwu Timur (Lutim), Sulawesi	
2	Mendes Minta Perda Sabtu, 22 Mei 2021 23:05	Jihan Khommasa - di Jakarta		Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigrasi	
3	Cempas M 3.6 Tarajek Sabtu, 22 Mei 2021 22:55	Tim detikcom - detikNews	Jakarta	Gempa berkekuatan magnitudo (M) 3,6 terjadi di Wangape, Sa	
4	Beri Benih hingga Co Sabtu, 22 Mei 2021 22:48	Nadhifa Sarah Amalia - Jakarta		Kementerian Pertanian (Kementan) serahkan bantuan benih hor	
5	MUI Setuju Saran X Sabtu, 22 Mei 2021 22:41	Tim detikcom - detikNews	Jakarta	Majelis Ulama Indonesia (MUI) meminta agar Pemerintah Indon	
6	Dialih Negeri? Pemas Sabtu, 22 Mei 2021 22:31	Tim detikcom - detikNews	Jakarta	Berkas kesadaran teknologi, pengguna yang memabrak lai pe	
7	Mendes Jadikan 2 Di Sabtu, 22 Mei 2021 22:01	Inkarna Putri - detikNews	Jakarta	Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigrasi	
8	PKP Bantah OTT Bu Sabtu, 22 Mei 2021 21:58	Mabus Albans - detikNews	Jakarta	Direktur Sosialisasi dan Kampanye Anti-Korupsi Komisi Pember	
9	Jajak Dostor Asal Su Sabtu, 22 Mei 2021 21:41	Tim detikcom - detikNews	Jakarta	Polisi Sumatera Utara (Sumut) menangkap empat orang tem	
10	Cegah Lonjakan CO2 Sabtu, 22 Mei 2021 21:40	Enika Dyah - detikNews	Jakarta	Satgas Penanganan COVID-19 meminta pemerintah kabupaten	
11	Lisur Lebaran, Vaku Sabtu, 22 Mei 2021 21:26	Nadhifa Sarah Amalia - Jakarta		Walaupun lama tetap dilaksanakan meski saat Idul Fitri, Vaku	
12	Saling Serang Panti Sabtu, 22 Mei 2021 21:01	Tim detikcom - detikNews	Jakarta	Tenaga Ahli Utama Kantor Staf Presiden (KSP) Ali Mochtar Ng	
13	Tak Dizinkan, Puku Sabtu, 22 Mei 2021 20:55	Adhyasta Digtarata - di Jakarta		Beberapa organisasi kemasyarakatan sipil bersama mahasiswa	
14	Kisah Jeda Tia, Puku Sabtu, 22 Mei 2021 20:50	Yogi Ernes - detikNews	Sering	Kasus penganiayaan anak yang dilakukan ayahnya sendiri di S	
15	Tekan COVID-19, Sa Sabtu, 22 Mei 2021 20:30	Inkarna Putri - detikNews	Jakarta	Juru Bicara Satgas Penanganan COVID-19 Prof Wiku Adisasmi	
16	Polisi & Satgas Diser Sabtu, 22 Mei 2021 20:28	Enika Dyah - detikNews	Jakarta	Ghribankambmas Desa Midang Poltek Gununggan, Apda Dew	
17	Sempat Hiang Beri Sabtu, 22 Mei 2021 20:16	Prima Syahbana - detikPalenbang		Empat gadis di Palenbang, Sumatera Selatan, yang dilaju	
18	Jl Juanda Belum Beri Sabtu, 22 Mei 2021 20:16	Adhyasta Digtarata - di Depok		Sejumlah pohon tumbang di beberapa titik di Jl H Juanda, Dep	
19	Detik-detik Mobil Por Sabtu, 22 Mei 2021 19:45	Yogi Ernes - detikNews	Jakarta	Sebuah mobil sport merek Porsche hangus terbakar di Kelapa C	

Gambar 8. Dataframe

3. Hasil dan Pembahasan

Metode pengumpulan data dengan *web scraping*, khususnya dari situs berita seperti `detik.com`, memberikan keuntungan dalam pengambilan informasi secara otomatis. Proses ini dilakukan dengan langkah-langkah yang terorganisir menggunakan *Jupyter Notebook* dan beberapa *library* penting, seperti *Beautiful Soup* dan *Requests*.

a) Instalasi Library:

Proses dimulai dengan menginstal library yang diperlukan, seperti *Beautiful Soup* dan *Requests*, untuk mempermudah proses *web scraping* di *Jupyter Notebook*.

b) Ekstraksi Isi Halaman Web:

- Melalui inspeksi halaman *web* `detik.com` menggunakan *Chrome DevTools*, penulis menganalisis struktur HTML untuk menentukan elemen-elemen yang perlu diambil. Ini mencakup judul, penulis, tanggal, dan isi berita.
- Hasil inspeksi digunakan sebagai panduan untuk menentukan *tag* dan *class* yang akan diekstraksi.

c) Parsing Halaman Web Menggunakan Beautiful Soup:

- Melalui proses *parsing*, library *Beautiful Soup* digunakan untuk mengekstrak informasi dari halaman *web* `detik.com`.
- Penentuan elemen-elemen seperti judul, penulis, tanggal, dan isi berita dilakukan berdasarkan struktur HTML yang telah diinspeksi.
- Pembersihan konten dilakukan, terutama pada elemen tabel di dalam isi berita, untuk mendapatkan teks yang bersih.

d) Mengambil Index Daftar Berita Pertanggal:

- Setelah berhasil mengambil data per berita, langkah selanjutnya adalah mengambil seluruh berita berdasarkan tanggal.

- Daftar berita pada tanggal tertentu terdapat pada halaman indeks yang memiliki parameter tanggal dan nomor halaman.
- Mengambil URL dari daftar berita untuk kemudian mengambil detail berita menggunakan *source code* yang disediakan.

e) Mengubah Dataset Menjadi Excel:

- Hasil data yang berhasil diambil dapat diubah menjadi format *Excel* menggunakan *library Pandas*.
- Fungsi `to_excel` pada *Pandas DataFrame* digunakan untuk menyimpan data ke dalam file Excel.

Dengan langkah-langkah ini, penulis berhasil membuat sebuah robot yang dapat mengumpulkan data berita dari `detik.com` secara otomatis. Proses ini memanfaatkan teknik *web scraping* dan *Python library* untuk mengambil, membersihkan, dan menyimpan data dengan efisien. Selanjutnya, hasil data dapat digunakan untuk analisis lebih lanjut atau tujuan tertentu, sesuai kebutuhan pengguna.

4. Kesimpulan

Penerapan teknik *web scraping* ini memberikan kemampuan untuk mengambil data besar-besaran dari `detik.com` tanpa harus melibatkan upaya manual yang besar. Namun, penting untuk diingat bahwa etika dan hukum penggunaan *web scraping* harus tetap diperhatikan, serta pemahaman terhadap aturan dan kebijakan situs *web* yang diambil datanya. Keseluruhan, penerapan *web scraping* membuka mendukung analisis lebih lanjut, dan dapat diaplikasikan dalam berbagai konteks tergantung pada kebutuhan pengguna.

Saya ingin menyampaikan rasa terima kasih yang mendalam atas dukungan dan kontribusi yang berharga dalam penulisan artikel ini. Proyek pengumpulan data dengan metode *web scraping* dari `detik.com` tidak mungkin terealisasi tanpa bantuan dan inspirasi dari berbagai pihak.

Daftar Pustaka

- [1] Zuo, Z. H., & Xie, Y. M. (2015). A simple and compact Python code for complex 3D topology optimization. *Advances in Engineering Software*, 85, 1-11.
- [2] Boulle, A., & Kieffer, J. (2019). High-performance Python for crystallographic computing. *Journal of Applied Crystallography*, 52(4), 882-897.
- [3] Pereira, F., Azevedo, F., Carvalho, A., Ribeiro, G. F., Budde, M. W., & Johansson, B. (2015). Pydna: a simulation and documentation tool for DNA assembly strategies using python. *BMC bioinformatics*, 16, 1-10.
- [4] Nasr, M., Giroux, B., & Dupuis, J. C. (2022). Python package for 3D joint hypocenter-velocity inversion on tetrahedral meshes: Parallel implementation and practical considerations. *Computational Geosciences*, 1-25.
- [5] Croce, T., Girardi, M., Gurioli, G., Padovani, C., & Pellegrini, D. (2023, August). Towards a Cloud-Based Platform for Structural Health Monitoring: Implementation and Numerical Issues. In *International Conference on*

Experimental Vibration Analysis for Civil Engineering Structures (pp. 610-619). Cham: Springer Nature Switzerland.