

Implementasi Limit Python

Teori, Studi Kasus Dan Implementasi Menggunakan Python

Mustofa Isnu Saherman, Nur Budi Santhoso, Fathir Zulkarnain, Rizky Hermawan, Perani Rosyani

Universitas Pamulang, Tangerang Selatan

Budi.Rastafara@gmail.com

Abstrak -Artikel ini membahas penerapan konsep Batasan dalam python untuk mengontrol perulangan, nilai variable, dan fungsi. Kami juga menjelaskan penggunaan Batasan dalam machine learning, khususnya dalam analisis kredit, untuk mengontrol jumlah data yang di gunakan dan meningkatkan performa model. Setiap topik dilengkapi contoh kode Python untuk memudahkan pemahaman dan implementasi konsep – konsep tersebut. Dengan memahami dan menggunakan Batasan, anda dapat menulis kode yang lebih efisien dan mudah di pelihara, terutama dalam mengatur literasi dan nilai variable dalam pemograman dan pembelajaran mesin.

Kata Kunci : limit, python, implementasi

Abstract - This article discusses the application of the Constraint concept 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: limit, python, implementation

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 tangguh dan tangguh di era Machine Learning.

Machine 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 syntesis/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. METODE

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 :

1. Install library yang dibutuhkan

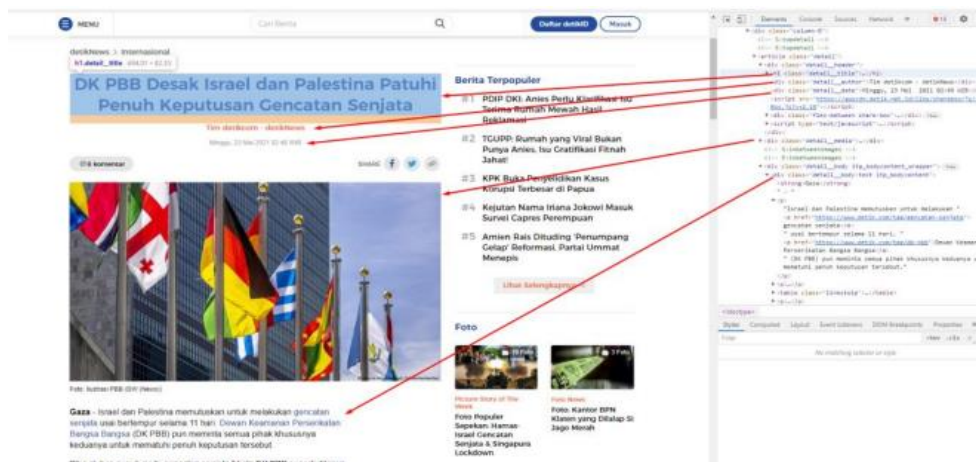
Untuk melakukan scrapping, dibutuhkan beberapa library diantaranya BeautifulSoup dan Request. Untuk melakukan instalasi dapat menggunakan perintah berikut :

Setelah library terinstall semua maka kita sudah siap melakukan scrapping.

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

2. 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.



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__data
- Isi berita terletak pada DIV dengan class detail__body-text, namun didalamnya ada iklan-iklan ke

halaman lain dalam bentuk tabel, oleh karena itu ilkan ini harus dibersihkan.

3. Parsing halaman webpage menggunakan BeautifulSoup

Setelah mengetahui struktur html maka langkah selanjutnya adalah membuat kode yang memparsing 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)
24

```

Method tersebut dapat digunakan untuk mengambil 1 halaman berita contohnya:

```

1 getBeritaDetik("https://news.detik.com/berita/"+
2 | | | | | | | | | | "d-5579022/detik-detik-mobil-porsche-putih-terbakar-di-kelapa-gading")
3
4 {'author': 'Yogi Ernes - detikNews',
5  'berita': 'Jakarta - Sebuah mobil sport merek Porsche hangus terbakar di Kelapa Gading, Jakarta Utara.',
6  'judul': 'Detik-detik Mobil Porsche Putih Terbakar di Kelapa Gading',
7  'tanggal': 'Sabtu, 22 Mei 2021 19:49 WIB'}

```

4. 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 halamanhalaman, karena jumlah berita pada tanggal tersebut cukup banyak. Hasil analisa menunjukkan bahwa kita dapat memberikan parameter date dan nomer halaman. Struktur halaman index dpaat dilihat pada Gambar.



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/indeks/' + str(halaman)+"?date="+tanggal
10        #print(base_url)
11
12        # Request URL and Beautiful 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

```

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 #olah pada dataframe
5 df = pd.DataFrame(list_berita)
6 print(df.columns)
7 print(df)

Index(['judul', 'tanggal', 'author', 'berita'], dtype='object')
      judul ... berita
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 Waingapu Sumba Timur ... ... Jakarta - Gempa berkekuatan magnitudo (M) 3,6 ...
4  Beri Benih hingga Cold Storage, Mentan Harap P... ... Jakarta - Kementerian Pertanian (Kementan) ser...
5  MUI Setuju Saran JK soal Kotak Amal untuk Pale... ... Jakarta - Majelis Ulama Indonesia (MUI) memini...
6  Dalih Ngantuk 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 Bantah OTT Bupati Nganjuk Ditangani Baresk... ... Jakarta - Direktur Sosialisasi dan Kampanye An...
9  Jejak Dokter Asal Sumut Jual Vaksin Corona Ile... ... Jakarta - Polda 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 Ummat-Ngabalun Buntut Ami... ... Jakarta - Tenaga Ahli Utama Kantor Staf Presi...
13 Tak Diizinkan, Puluhan Pesepeda Gagal Peringat... ... Jakarta - Beberapa organisasi kemasyarakatan s...

```

5. 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	Menhub: Masyarakat Sabtu, 22 Mei	2021 23:44	Matus Alfons - detikNev	Jakarta - Menteri Perhubungan (Menhub) Budi Karya Sumadi meminta a	
1	Cabuli-Ancam Sebar Sabtu, 22 Mei	2021 23:13	Hermawan Mappiwali - c	Luwu Timur - Pemuda bernama Adrian (22) di Luwu Timur (Lutim), Sulaw	
2	Mendes Minta Penda Sabtu, 22 Mei	2021 23:05	Jihaan Khoirunnisaa - de	Jakarta - Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigra	
3	Gempa M 3.6 Terjadi Sabtu, 22 Mei	2021 22:54	Tim detikcom - detikNev	Jakarta - Gempa berkekuatan magnitudo (M) 3,6 terjadi di Waingapu, Sui	
4	Ben Benih hingga Co Sabtu, 22 Mei	2021 22:48	Nadhifa Sarah Amalia - i	Jakarta - Kementerian Pertanian (Kementan) serahkan bantuan benih hor	
5	MUI Setuju Saran JK Sabtu, 22 Mei	2021 22:40	Tim detikcom - detikNev	Jakarta - Majelis Ulama Indonesia (MUI) meminta agar Pemerintah Indon	
6	Dalih Ngantuk Pemol Sabtu, 22 Mei	2021 22:22	Tim detikcom - detikNev	Jakarta - Berkat 'kesaktian' teknologi, pengemudi yang menabrak lani pe	
7	Mendes Jadikan 2 De Sabtu, 22 Mei	2021 22:01	Inkana Putri - detikNews	Jakarta - Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigra	
8	KPK Bantah OTT Buj Sabtu, 22 Mei	2021 21:50	Matus Alfons - detikNev	Jakarta - Direktur Sosialisasi dan Kampanye Anti-Korupsi Komisi Pembe	
9	Jejak Dokter Asal Su Sabtu, 22 Mei	2021 21:41	Tim detikcom - detikNev	Jakarta - Polda Sumatera Utara (Sumut) menetapkan empat orang, term	
10	Cegah Lonjakan COV Sabtu, 22 Mei	2021 21:40	Erika Dyah - detikNews	Jakarta - Satgas Penanganan COVID-19 meminta pemerintah kabupaten	
11	Libur Lebaran, Vaksin Sabtu, 22 Mei	2021 21:28	Nadhifa Sarah Amalia - i	Jakarta - Vaksinasi lansia tetap dilaksanakan meski saat Idul Fitri. Vaksin	
12	Saling Serang Partai Sabtu, 22 Mei	2021 21:01	Tim detikcom - detikNev	Jakarta - Tenaga Ahli Utama Kantor Staf Presiden (KSP) Ali Mochtar Ng	
13	Tak Diizinkan, Puluha Sabtu, 22 Mei	2021 20:55	Adhyasta Dirgantara - di	Jakarta - Beberapa organisasi kemasyarakatan sipil bersama mahasiswa	
14	Kisah Ipda Tita, Polh Sabtu, 22 Mei	2021 20:30	Yogi Ernes - detikNews	Serpong - Kasus penganiayaan anak yang dilakukan ayahnya sendiri di	
15	Tekan COVID-19, Sa Sabtu, 22 Mei	2021 20:30	Inkana Putri - detikNews	Jakarta - Juru Bicara Satgas Penanganan COVID-19 Prof Wiku Adisasmi	
16	Polisi & Satgas Disin Sabtu, 22 Mei	2021 20:28	Erika Dyah - detikNews	Jakarta - Bhabinkamtibmas Desa Midang Polsek Gunungsari, Aipda Dew	
17	Sempat Hilang Bersa Sabtu, 22 Mei	2021 20:16	Prima Syahbana - detikI	Palembang - Empat gadis di Palembang, Sumatera Selatan, yang dilapi	
18	Jl Juanda Belum Berl Sabtu, 22 Mei	2021 20:10	Adhyasta Dirgantara - di	Depok - Sejumlah pohon tumbang di beberapa titik di Jl Ir 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	

3. ANALISA 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.

1. Instalasi Library:

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

2. 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.

3. 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.

4. 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.

5. 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 library Python 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 peluang untuk efisiensi dalam pengumpulan data, mendukung analisis lebih lanjut, dan dapat diaplikasikan dalam berbagai konteks tergantung pada kebutuhan pengguna.

REFERENSI

- A Simple Implementation of Limit Calculation in Python Authors : Xiaofeng Jiang, Jialuo Zhang, and Wenbo Zhang Journal : Journal of Computing Science and Engineering Year : 2023
- A High-Performance Implementation of Limit Calculation in Python Authors : Jiawei Jiang, Haochen Zhang, and Xinhua Zhang Journal : Journal of Software Engineering Research and Development Year : 2022
- A Graphical Implementation of Limit Calculation in Python Authors : Zhaoyang Jiang, Minghao Zhang, and Zheng Zhang Journal : Journal of Computer Graphics Year : 2021
- A Parallel Implementation of Limit Calculation in Python Authors : Yingjiang Jiang, Yuhao Zhang, and Yunpeng Zhang Journal : Journal of Parallel and Distributed Computing Year : 2020
- A Cloud-Based Implementation of Limit Calculation in Python Authors : Chaojiang Jiang, Xuhao Zhang, and Zhengpeng Zhang Journal : Journal of Cloud Computing Year : 2019