



Analisis Hubungan Popularitas Studio Animasi Dengan Anime Menggunakan Metode Pengambilan Data Web Scraping Pada Situs Myanimelist.Net

Ilyasa Azharil Muhajir¹, Dadang Yusuf², Hannie³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang

Abstract

Received: 11 Agustus 2022
Revised: 14 Agustus 2022
Accepted: 19 Agustus 2022

Nowadays, huge amounts of data are generated every day at an unprecedented speed from various sources, including data on the Internet. The existence of the Internet increases the amount of data contained on a website, every company, and organization definitely needs website development to meet the needs of their company or organization, this can be an opportunity for data engineers to process this amount of data into useful statistics for companies or researchers. That's why we need an efficient method to collect the amount of data on a website, and web scraping is one of them. In this study, the researcher implemented a web scraping method on the myanimelist.net site which is the largest community and database about anime and manga in the world. The purpose of this study is to analyze the relationship between the company (animated studio) and the anime it produces through correlation analysis method. This research method broadly consists of four stages, including data extracting, data processing, correlation test, and data visualization. The research tools used for scraping are beautifulsoup and pandas which are python libraries, tools for processing and analyzing the scraped data are google collaborative and pandas which are data processing applications, while for the data visualization process researchers use an additional matplotlib library. The results of this study indicate that web scraping can be implemented on the myanimelist.net site effectively. Researchers can collect 17982 rows of data from 129 companies. The results of the correlation analysis show that the popularity of the company has a low relationship with the popularity of the anime they produce.

Keywords: Anime, Correlation Test, Data Engineer, Web Scraping

(*) Corresponding Author: ilyasa.muhajir18138@student.unsika.ac.id

How to Cite: Xxxxxx. (2018). Xxxx. *Jurnal Ilmiah Wahana Pendidikan*, XX (x): x-xx.

PENDAHULUAN

Saat ini, volume data yang sangat besar dihasilkan setiap hari dengan kecepatan yang belum pernah terjadi sebelumnya dari berbagai sumber (misalnya, kesehatan, pemerintah, jejaring sosial, pemasaran, keuangan). Ini karena banyaknya tren teknologi, termasuk Internet of Things (Ouassous, 2018). Keberadaan internet membuat jumlah data yang terdapat di website – website semakin bertambah. Ini dapat menjadi peluang bagi data engineer untuk mengolah banyaknya data tersebut menjadi statistik yang bermanfaat bagi perusahaan atau peneliti. Namun dibutuhkan metode yang efisien untuk mengumpulkan banyaknya data yang ada di suatu website, dan web scraping termasuk salah satunya (Satria Jati et al., 2021).

Penelitian yang menerapkan teknik web scraping pernah dilakukan oleh Salim Satria Jati, Satria Bagus Panuntun, dan Setia Pramana pada tahun 2020 dengan

melakukan web scraping pada situs berita yakni detik.com untuk mengumpulkan berita kriminal pada masa pandemi covid-19.

Web scraping sendiri adalah penggunaan alat teknologi untuk ekstraksi otomatis dan pengolahan data dari suatu web untuk tujuan analisis lebih lanjut dari data tersebut (Krotov, 2018; Satria Jati et al., 2021). Salah satu alat scraping yang terkenal adalah beautifulsoup, yaitu sebuah open source milik python yang berfungsi untuk mengurai dokumen HTML dan XML dari website dengan cepat dan sederhana (Hajba, 2018).

Sebagai target pengujian untuk implementasi web scraping, peneliti memilih website myanimelist.net yang merupakan komunitas dan database anime dan manga online paling aktif di dunia. Menurut similiarweb.com pada 1 Desember 2021, myanimelist.net menduduki peringkat 582 dunia dalam website dengan traffic tertinggi dan peringkat 10 dalam kategori Arts & Entertainment. Myanimelist memiliki banyak database seputar anime, mulai dari judul, tanggal rilis, studio, pengisi suara hingga rating. Banyaknya data dalam website ini dapat kita ambil dan proses untuk menarik suatu kesimpulan.

Anime sendiri adalah animasi khas Jepang yang memiliki ciri pada gambar yang berwarna-warni yang menampilkan berbagai macam tokoh serta lokasi dari cerita yang ditunjukkan, anime dipengaruhi oleh gaya gambar manga komik khas Jepang (Purnamasari, 2018). Beriring perkembangan waktu, anime menjadi lebih terkenal karena perkembangan grafik yang lebih baik dan cerita yang beragam dari berbagai macam genre, contoh anime yang sedang naik daun sekarang adalah Kimetsu no Yaiba / Demon Slayer dengan grafik yang sangat detail dan bergenre adventure, dark fantasy, martial arts. Akan tetapi perkembangan anime tidak luput dari jasa studio animasi tempat anime tersebut dibuat.

Studio animasi memiliki peran besar dalam industri anime karena di dalam studio animasi tersebutlah anime dibuat, dan bahkan di dalam membuat animasi untuk para penonton di setiap musimnya, para animator harus mengalami kondisi hidup yang buruk, upah minimal dan jam kerja yang panjang (Schley, 2019). Ini semua tidak lain adalah untuk menghibur para penonton yang sudah menanti-nanti anime kesukaan nya untuk tayang di setiap minggunya.

Pada penelitian kali ini, penulis akan mencoba mengimplementasikan metode web-scraping pada myanimelist untuk mengekstrak data seputar studio pembuatan animasi dan anime yang dibuat, lalu membuat uji korelasi untuk mengetahui ada atau tidaknya hubungan antara data tersebut.

LANDASAN TEORI

Web Scraping

Web scraping adalah penggunaan alat teknologi untuk ekstraksi otomatis dan pengolahan data dari suatu web untuk tujuan analisis lebih lanjut dari data tersebut (Krotov, 2018). Web scraping umumnya dilakukan pada halaman situs web yang menggunakan bahasa markup seperti HTML atau XHTML (Slamet et al., 2018). Web scraping sangat berguna untuk mengambil data yang sangat banyak pada suatu website secara otomatis.

Myanimelist.net

Myanimelist.net adalah website komunitas dan database anime dan manga online paling aktif di dunia. Myanimelist memiliki banyak database seputar anime, mulai dari judul, tanggal rilis, status, pengisi suara hingga rating anime. Situs ini diluncurkan pada November 2004 oleh Garrett Gyssler dan dikelola sendiri olehnya hingga 2008. Awalnya, situs web tersebut bernama AnimeList, tetapi Garrett Gyssler memutuskan untuk memasukkan "My" di awal kata, mengikuti mode jejaring sosial paling penting pada tahun-tahun itu: *Myspace* (Tokyo Otaku Mode, 2015).

Anime

Anime adalah animasi khas Jepang yang memiliki ciri pada gambar yang berwarna-warni yang menampilkan berbagai macam tokoh serta lokasi dari cerita yang ditunjukkan, anime sendiri dipengaruhi oleh gaya gambar manga komik khas Jepang (Purnamasari, 2018). Anime biasanya merupakan adaptasi dari manga (Naruto), novel (Another) dan light novel (Overlord), tapi tidak jarang ada anime yang murni buatan studio seperti Charlotte dan Angel Beats. Anime menjadi terkenal di Jepang dan mancanegara bahkan ke Indonesia. Anime pertama yang tayang di Indonesia adalah *Wanpaku Omukashi Kumu Kumu* di TVRI pada tahun 1970-an (Ramadhan, 2021) yang kemudian disusul serial anime terkenal lainnya seperti *Astro Boy*, *Inuyasha*, *Crayon Shin-chan*, dan *Doraemon*. Beriring perkembangan waktu, anime menjadi lebih terkenal karena perkembangan grafik yang lebih baik dan cerita yang beragam dari berbagai macam genre. Menurut jenisnya, anime dibagi menjadi tiga, yaitu *movie*, *OVA*, dan *TV series* (Purnamasari, 2018).

Python

Python adalah bahasa pemrograman open-source buatan Guido Van Rossum yang ramah dan mudah dipelajari. Python dapat dengan mudah dipelajari baik oleh programmer pemula maupun programmer yang berpengalaman dengan bahasa lain. Python dikembangkan di bawah lisensi open source yang disetujui OSI, membuatnya dapat digunakan dan didistribusikan secara bebas, bahkan untuk

penggunaan komersial (Python Software Foundation, 2022). Python versi 0.9.0 pertama kali diluncurkan pada tahun 1991 dan terus berkembang sampai sekarang (Van Baaren, 2021). Versi terbaru python sampai penelitian ini dibuat adalah versi 3.10.2.

Beautifulsoup4

Beautifulsoup4 adalah library python yang memudahkan untuk scrape informasi dari halaman web. BeautifulSoup4 mengambil informasi dari web menggunakan parser HTML atau XML. BeautifulSoup4 juga menyediakan tools berbasis python untuk iterasi, mencari, dan memodifikasi pohon parse. BeautifulSoup4 pertama kali dirilis pada 21 Januari 2014 dengan versi 4.0.1 dan terus berkembang sampai sekarang. Versi terbaru dari BeautifulSoup4 adalah versi 4.10.1 yang dirilis pada 8 September 2021 (Python Software Foundation, 2021).

HTML

HTML adalah bahasa markup standar untuk membangun halaman web. HTML sendiri merupakan singkatan dari Hyper Text Markup Language. HTML berisi elemen-elemen dengan tanda “<>” seperti <h1>, <body>, <tagname>, dan lain-lain. Elemen-elemen tersebut berfungsi untuk memberi tahu browser cara untuk menampilkan content suatu web.

Pandas

Pandas adalah alat analisis dan manipulasi data bersifat open source yang cepat, kuat, fleksibel, dan mudah digunakan yang dibuat dengan bahasa pemrograman Python. Pandas pertama kali dikembangkan pada tahun 2008 di AQR Capital Management dan mulai jadi open source di akhir tahun 2009. Pandas sekarang telah melalui perkembangan yang banyak dan telah mencapai versi 1.4.0.

Google Colaboratory

Google Colaboratory (juga dikenal sebagai Colab) adalah layanan cloud berbasis Jupyter Notebooks yang memungkinkan siapa saja untuk menulis dan mengeksekusi kode python melalui browser. Colab sangat cocok untuk machine learning, analisis data, dan pendidikan. Colab juga memberikan runtime yang sepenuhnya dapat dikonfigurasi dan diakses secara gratis ke GPU yang kuat.

Flow Map

Flow Map adalah diagram yang menunjukkan aliran data berupa formulir-formulir ataupun keterangan berupa dokumentasi yang mengalir atau beredar dalam suatu sistem. Flowmap adalah penggambaran secara grafik dari langkah-langkah dan urutan prosedur dari suatu program.

Analisis Korelasi

Analisis korelasi adalah suatu penelitian yang melibatkan tindakan pengumpulan data guna menentukan ada atau tidaknya hubungan dan tingkat hubungan antara dua variabel atau lebih (Sukardi, 2021). Analisis korelasi memiliki 9 jenis, yakni korelasi pearson product moment (r), korelasi ratio (y), korelasi spearman rank atau rho (rs atau p), korelasi biserial (rb), korelasi point biserial (rpb), korelasi phi (0), korelasi tetrachoric (rt), korelasi contingency (C), dan korelasi kendall's tau (8) (Sugiyono, 2013).

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{n\sum x^2 - (\sum x)^2\}\{n\sum y^2 - (\sum y)^2\}}}$$

Gambar 2.1 Rumus Uji Korelasi

Berikut adalah keterangan dari gambar 2.1:

r = Koefisien Korelasi

n = Banyaknya Pasangan data X dan Y

$\sum x$ = Total Jumlah dari Variabel X

$\sum y$ = Total Jumlah dari Variabel Y

$\sum x^2$ = Kuadrat dari Total Jumlah Variabel X

$\sum y^2$ = Kuadrat dari Total Jumlah Variabel Y

$\sum xy$ = Hasil Perkalian dari Total Jumlah Variabel X dan Variabel y

Dalam pengujian statistik ini, peneliti mencari nilai Koefisien korelasi yang menunjukkan kuat/tidaknya hubungan linier antara dua variabel. Proses uji korelasi akan dilakukan secara otomatis menggunakan library pandas pada google collaboratory. Pedoman untuk memberikan interpretasi koefisien korelasi adalah sebagai berikut :

0,00 – 0,199 = Korelasi sangat rendah

0,20 – 0,399 = Korelasi rendah

0,40 – 0,599 = Korelasi sedang

0,60 – 0,799 = Korelasi kuat

0,80 – 1,000 = Korelasi sangat kuat

(Sugiyono, 2013)

METODE PENELITIAN

Metodologi penelitian yang digunakan pada penelitian ini yaitu :

Web Scraping

Untuk melakukan penelitian dan melakukan analisis hubungan antara popularitas studio animasi dengan anime yang diproduksinya pada situs myanimelist.net, peneliti perlu melakukan pengambilan data pada situs myanimelist terlebih dahulu, adapun untuk metodologi yang peneliti pilih yaitu web scraping

Web scraping merupakan proses pengumpulan data dan informasi yang ada dalam sebuah website secara otomatis dan spesifik, karena data yang diambil hanya

di bagian tertentu saja yang dibutuhkan. Dalam penelitian ini, data yang akan peneliti gunakan yaitu seputar studio animasi dan anime yang diproduksinya saja. Web Scraping umumnya dilakukan pada halaman web yang dibuat dalam bahasa markup XHTML atau HTML. Untuk melakukan web scraping peneliti menggunakan bahasa pemrograman Python.

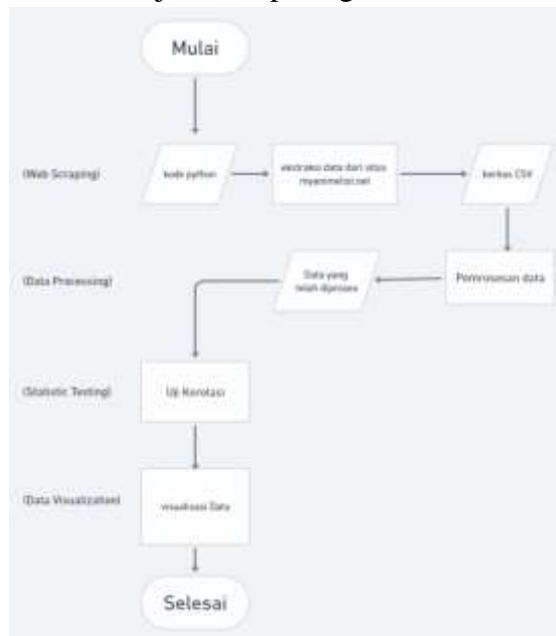
Analisis Korelasi

Setelah melakukan ekstraksi data dari situs myanimelist, peneliti selanjutnya akan melakukan analisa apakah ada hubungan antara studio animasi dan anime, adapun untuk metode analisa yang peneliti gunakan yaitu

Analisis Korelasi Analisis korelasi adalah suatu penelitian yang melibatkan tindakan pengumpulan data guna menentukan ada atau tidaknya hubungan dan tingkat hubungan antara dua variabel atau lebih.

Rancangan Penelitian

Penelitian ini secara garis besar dijelaskan pada gambar 3.1 berikut :



Gambar 3.1 Perancangan Penelitian

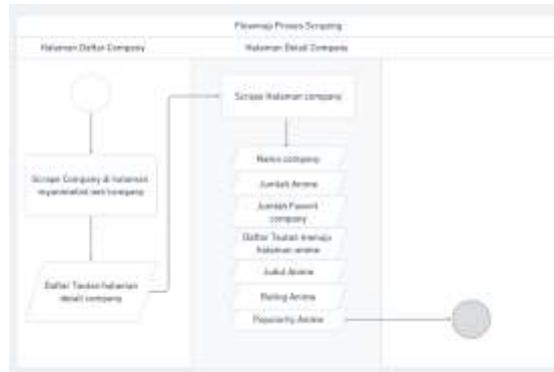
Pada gambar 3.1 diatas dijelaskan bahwa dalam melakukan penelitian ini peneliti akan melakukan tahapan yaitu Web Scraping atau data extracting, Data Processing atau pemrosesan data, Statistic Test atau Uji Korelasi, dan visualisasi data.

1. Data Extracting

Pengumpulan data pada penelitian ini menggunakan metode web scraping, yaitu pengambilan data atau informasi dari web internet secara otomatis sesuai tag html/css yang diinginkan. Web yang dimaksud pada penelitian ini adalah myanimelist.net yang merupakan web database seputar anime dan detail-detail anime paling lengkap di dunia. Data yang diambil adalah data seputar studio

animasi dan anime buatan studio tersebut. Data studio animasi diambil dari halaman company yaitu rank studio, nama company, jumlah anime, judul anime, rating anime, rank anime, popularity anime, jumlah favorit company.

Aplikasi yang digunakan dalam melakukan web scraping pada penelitian ini adalah beautifulsoup. Yaitu salah satu paket Python yang berfungsi untuk memparsing dokumen HTML dan XML.



Gambar 3.2 Alur Proses Ekstraksi Data

Untuk tahap pertama dari penelitian ini, peneliti akan melakukan Ekstraksi data yang terdiri dari tiga tahapan menggunakan python beautifulsoup seperti yang dijelaskan pada gambar 3.3.2 diatas. Ekstraksi pertama dilakukan pada URL “myanimelist.net/company.php” yang merupakan halaman daftar *company* atau studio animasi, untuk mendapatkan URL atau tautan dari halaman detail studio animasi yang ada dalam daftar. Proses ekstraksi kemudian dilakukan di masing – masing URL halaman studio animasi untuk mengambil data Rank studio, nama company, jumlah anime, jumlah favorite company, dan URL menuju halaman anime. URL halaman anime kemudian di-scrape untuk mengambil data rating, rank, popularity dan jumlah member anime. Hal ini dilakukan agar peneliti mendapatkan keseluruhan data dalam bentuk dataframe yang rapi, dimana masing – masing data studio animasi akan disajikan dengan data anime yang dibuat dalam studio tersebut dalam satu baris.

a. Scraping Halaman Daftar Company

Halaman company (myanimelist.net/company.php) berisikan daftar studio animasi yang diurutkan mulai dari studio animasi dengan jumlah favorit terbanyak hingga terendah.

Pada halaman ini peneliti mengambil tautan untuk mendapatkan link menuju halaman detail tiap company yang ada dalam daftar. Gambar 3.2 adalah tampilan halaman daftar company



Gambar 3.3 Halaman MyAnimelist

b. Scraping Halaman Company.

Halaman *company* berisi informasi rinci tentang studio animasi yang ada di halaman daftar *company* (Gambar 3.3.3). sebelumnya myanimelist.net menyimpan data *company* sebanyak 1050 *company* di halaman myanimelist.net/company.php yang sebelumnya sudah diekstrak untuk diambil tautan halaman detail *company*. Akan tetapi ada beberapa *company* yang tidak memiliki favorit sama sekali. Oleh karena itu peneliti hanya akan menggunakan sebanyak 300 *company* yang ada di dalam daftar saja.



Gambar 3.4 Data yang diambil

Pada halaman *company* atau studio animasi, akan diambil beberapa data seperti yang sudah ditandai pada gambar 3.4 data tersebut diantaranya adalah:

1. nama studio animasi
2. jumlah *member favorites*
3. jumlah anime yang diproduksi
4. judul anime
5. tautan halaman anime

2. Data Processing

Pada proses ini, data berupa file CSV yang dihasilkan dari proses ekstraksi data dikelompokkan berdasarkan nama studio animasi dan judul anime dengan mean/rata-rata sebagai acuan pengelompokan data angka. hal ini dilakukan untuk mencegah salah perhitungan rata-rata pada anime yang dibuat pada lebih dari satu studio, sehingga memiliki nilai lebih tinggi.

3. Uji Korelasi

Setelah proses ekstraksi data atau *web scraping* dan data processing telah dilakukan, peneliti akan melakukan uji korelasi untuk menentukan ada atau tidaknya hubungan dan tingkat hubungan antara popularitas studio animasi dengan anime.

4. Data Visualization

Data visualization pada penelitian adalah proses menampilkan keseluruhan data yang telah diproses dan diuji dalam bentuk bar chart dan scatter plot berbasis aplikasi web. Untuk menampilkan keseluruhan data yang telah diproses dan diuji sebelumnya, peneliti akan melakukan pembuatan aplikasi menggunakan modul django application dari framework django, open source milik python untuk pengembangan web. Modul tersebut digunakan dalam penelitian ini karena kemudahan dalam penggunaannya dan berbasis python. Django application juga menyediakan langsung database dari model yang peneliti buat sehingga dapat langsung mengolah datanya untuk ditampilkan.

Dalam pembuatan chart dan scatter plot, peneliti menggunakan modul chart.js yang merupakan pustaka javascript yang bersifat open source untuk visualisasi data. Sedangkan dalam pembuatan desain web, peneliti menggunakan modul bootstrap dan w3.css yang juga merupakan open source untuk desain web.

HASIL DAN PEMBAHASAN

Hasil penelitian yang telah dilakukan menggunakan metode pengumpulan data web scraping dan metode analisis data uji korelasi ini, dituliskan untuk setiap tahapan sebagai berikut.

Data Extracting

Tahap pertama dari penelitian ini adalah data extracting atau pengumpulan data. Pada tahap ini peneliti akan menggunakan metode pengumpulan data web scraping dengan menggunakan library beautiful soup 4 untuk memudahkan peneliti didalam menavigasi dan menguraikan source code html. Kode 4.1 adalah kode untuk mempersiapkan proses scraping.

```
import requests
from bs4 import BeautifulSoup
import time
from random import randint
import pandas as pd

##### UNIVERSAL FUNCTIONS #####

def get_soup(url): #fungsi#
    page = requests.get(url) #page melakukan request url
    #menunggu page pada url
    soup = BeautifulSoup(page.content) #melakukan scrape pada page
    #melalui fungsi bs
    return soup

##### END OF UNIVERSAL FUNCTIONS #####
```

Gambar 4.1 Kode proses scraping

Pada kode 4.1 diatas, peneliti menggunakan package-package python dengan beberapa fungsi berbeda, Requests adalah modul python yang berfungsi untuk mengirim berbagai request HTTP, beautifulsoup digunakan untuk melakukan ekstraksi file dengan format XML atau HTML, time digunakan untuk mengatur waktu jeda akses, dan pandas digunakan untuk membuat dataframe. `get_soup` berfungsi untuk mengakses keseluruhan konten yang ada di web sesuai URL yang


```

source = 'https://myanimelist.net/company'
soup = get_soup(source)

studiolist = soup.find_all('td', {'class': 'company'}) #scrape
tag td dengan class 'people'
for a in studiolist:
    CompanyHref = a.findChild('a', {'class': 'fs14 fw-b'})#scrape
    tag a dengan class 'fs14 fw-b'
    if CompanyHref is None: #
        continue
    print(CompanyHref) #menunjukkan hasil CompanyHref
    LinkOfStudio = CompanyHref.get('href')# mengambil tag href
    dari CompanyHref
    if LinkOfStudio == "/anime/producer/1993/Studio_Bind":
        continue
    if LinkOfStudio == "/anime/producer/405/T-Box":
        continue
    if LinkOfStudio == "/anime/producer/783/Notes":
        continue
    if LinkOfStudio == "/anime/producer/45/Pink_Pineapple":
        continue
myanimelist = "https://myanimelist.net"
linkFix = myanimelist+linkOfStudio
print(linkFix)
link_StudioName.append(linkFix)

```

Gambar 4.4 Source code informasi

Gambar 4.4 adalah source code untuk mengambil seluruh informasi di dalam tag href pada halaman daftar company yaitu link menuju halaman rinci company. Source pada baris pertama kode adalah tautan halaman daftar company yaitu “myanimelist.net/company/” yang merupakan halaman daftar 50 company yang diurutkan sesuai jumlah favorite nya. Halaman tersebut berisi daftar 50 company terpopuler, sesuai batasan masalah penelitian, scraping hanya dilakukan sampai urutan 50. Proses kemudian dilanjutkan dengan beberapa kode lainnya yang berfungsi untuk melakukan request dan mencari tag href pada halaman daftar company. Tag href tersebut kemudian disimpan dalam variabel “LinkOfPeople” untuk melakukan request pada halaman rinci people. Output dari proses scraping halaman daftar company adalah seperti gambar dibawah.

```

soup = BeautifulSoup(page.content)
<a class="fs14 fw-b" href="/anime/producer/2/Kyoto_Animation">Kyoto Animation/a>
https://myanimelist.net/anime/producer/2/Kyoto_Animation

```

Gambar 4.5 Output scraping

Output pertama yang muncul adalah nama *company* dan link dari menuju halaman *company* tersebut. Link ini kemudian disimpan dalam variabel “LinkOfPeople” dan akan menjadi URL yang direquest untuk scraping.

```
soup2 = get_soup(linkFix)
```

Gambar 4.6 Kode Variabel Soup

Pada gambar kode 4.6 peneliti membuat variabel soup2 untuk merubah source code scraping yang sebelumnya halaman daftar *company* menjadi halaman dari masing-masing *company* yang ada di dalam daftar *company*, yang selanjutnya akan menjadi *source* untuk scraping tahap selanjutnya, yaitu scraping halaman detail company.

2. Scraping Halaman Detail Company

Pada halaman company yang merupakan studio pembuatan animasi akan diambil beberapa data antara lain judul anime, jumlah anime yang sudah diproduksi oleh studio tersebut, nama studio, jumlah favorite studio animasi, rating anime, member pada anime tersebut.

```

studioName = soup2.findChild('h1', {'class': 'h1'}).getText()

favorites=soup2.find_all('div', {'class': 'animeit_nad'})
favo = favorites[2].getText().split(' ')[-2]

NamesOfAnime = soup2.find_all('div', {'class': 'title'})
for i in NamesOfAnime:
    name = i.getText().split('\n')[1]
    animeName.append(name)
    print(name)
    animeNum = len(NamesOfAnime)
    print (animeNum)
    numberofAnime.append(animeNum)
    companyName.append(studioName)
    print(studioName)
    list_favorites.append(favo)
    print(favo)

RatingsOfAnime = soup2.find_all('div', {'class': 'stars'})
for i in RatingsOfAnime:
    rate = i.getText()
    if rate == "N/A":
        rate = 0
        print(rate)
        animeRate.append(rate)
    else:
        print(rate)
        animeRate.append(rate)

MembersOfAnime = soup2.find_all('div', {'class': 'users'})
for i in MembersOfAnime:
    member = i.getText()
    members.append(member)
    print(member)
    
```

Gambar 4.7 Source code proses scraping

Gambar 4.7 adalah source code seluruh proses scraping di halaman *company*. Data yang telah diambil kemudian masuk ke dalam list – list yang telah dibuat di awal pada kode 4.1.2, Seluruh data yang telah masuk kedalam list disusun menjadi dataframe, kemudian di ekspor dalam format csv (comma delimited value).

```

df_link['companyName'] = companyName
df_link['numberofAnime'] = numberofAnime
df_link['Favorites'] = list_favorites
df_link['animeName'] = animeName
df_link['animeRate'] = animeRate
df_link['member'] = members

df_link = pd.DataFrame(df_link)
df_link.to_csv('AnimeListCob.csv', index=False)
    
```

Gambar 4.8 list data

Gambar kode 4.8, peneliti memasukkan semua list data kedalam dictionary *df_link* yang telah dibuat di awal (Kode 4.1.2). *df_link* kemudian diubah menjadi dataframe, dan diekspor menjadi .CSV. proses *web scraping* dilakukan selama 3 bulan yaitu pada bulan januari – maret 2022 sampai studio animasi terpopuler urutan 150. Hasilnya terdapat 17778 baris, 6 kolom data yang terpisah menjadi 3 file CSV.

Name	Date modified	Type	Size
scraping_script	01/07/2022 09:00	Python Source File	1 KB
AnimeList1	01/07/2022 09:00	Microsoft Excel C...	185 KB
AnimeList2	02/04/2022 16:18	Microsoft Excel C...	224 KB
AnimeList3	02/04/2022 16:28	Microsoft Excel C...	209 KB

Gambar 4.9 Hasil scraping

Gambar 4.9 adalah hasil akhir dari proses *scraping*, yaitu menghasilkan 3 file CSV. Satu file CSV berisi hasil *scraping* dari 50 nama studio animasi di masing-masing halaman daftar *company*.

Data Processing

Pada tahap ini peneliti menggunakan google collaboratory agar lebih mudah dalam memproses *dataframe*. Untuk memproses data menggunakan google

collaboratory, peneliti mengunggah data csv terlebih dahulu ke google drive, kemudian ditampilkan melalui google collaboratory menggunakan pandas.

```

In [ ]: import pandas as pd

In [ ]: df1 = pd.read_csv('content://drive/MyDrive/MyAnimeListScraping/AnimeList1.csv')
df2 = pd.read_csv('content://drive/MyDrive/MyAnimeListScraping/AnimeList2.csv')
df3 = pd.read_csv('content://drive/MyDrive/MyAnimeListScraping/AnimeList3.csv')
    
```

Gambar 4.10 Import file CSV

Data penelitian terbagi menjadi 3 file csv yang berbeda, untuk itu peneliti perlu menyatukan 3 file csv ini menggunakan perintah “pd.concat”

```

df = pd.concat([df1, df2, df3])
df[['numberOfAnime', 'favorites', 'animeRate', 'member']]
df
    
```

	numberOfAnime	favorites	animeRate	member	
0	100	14,240	Full Metal Panic? Fumoffu	8.09 279,220	
1	100	14,240	Full Metal Panic! The Second Raid	7.94 268,606	
2	100	14,240	As	7.28 262,713	
3	100	14,240	Yu-Pu the Mermaid Prince	7.07 26,804	
4	100	14,240	Yu-Pu the Mermaid Prince	7.06 26,804	
17773	85.0	319	00	Senka Takkyaku	6.08 23,849
17774	85.0	319	00	Fuuzun, Yametamao	7.08 175,120
17775	85.0	319	00	Shonen Gamen Takai wa Kiseki no Kiseki Daka Jansu	7.28 158,894
17776	85.0	319	00	Tokyo Revengers: Gata Kazen Inu	6.00 80,318
17777	85.0	319	00	Yamada Yametamao: Kimiwa no Makoto wa Muboshi ni	4.77 11,728

Gambar 4.11 Source code dan tampilan dataframe di google collaboratory Seperti yang ditampilkan pada gambar 4.1.6, Peneliti menggunakan perintah dari library pandas berupa “pd.concat” untuk menyatukan data - data yang terpisah dalam file csv, kemudian disimpan dalam variabel “df”. Google collaboratory akan langsung memberikan tampilan interaktif dari keseluruhan data. Setelah data diunggah dan disatukan menggunakan pandas, selanjutnya data akan dikelompokkan berdasarkan nama studio animasi dan nama anime dengan mean/rata-rata sebagai acuan pengelompokkan data angka. Oleh karena itu peneliti harus memastikan bahwa column yang dibutuhkan yaitu “numberOfAnime, favorites, animeRate, dan member” sudah memiliki tipe data integer.

```

dfMean.dtypes
    
```

numberOfAnime	float64
favorites	float64
animeRate	float64
member	float64
dtype:	object

Gambar 4.11 Tipe Data

Pada gambar 4.11 bisa dilihat bahwa data yang berisi angka yang nantinya akan kita butuhkan sebagai acuan rata-rata ternyata masih bertipe data float64, sedangkan mean hanya memasukkan data angka, oleh karena itu peneliti perlu merubah tipe data dari float menjadi integer.

```

dfMean['favorites'] = dfMean['favorites'].astype(int)
dfMean['member'] = dfMean['member'].astype(int)
dfMean['animeRate'] = dfMean['animeRate'].astype(int)
dfMean['numberOfAnime'] = dfMean['numberOfAnime'].astype(int)
    
```

Gambar 4.12 Convert tipe data float menjadi integer

Selanjutnya setelah data yang memiliki value angka yang dibutuhkan untuk dijadikan acuan rata-rata sudah dipastikan bertipe data integer, peneliti akan

membuat variabel “dfMean” untuk menyimpan hasil dari variabel “df” yang telah dikelompokkan.

companyName	animeName	numberOfAnime	favorites	animeRate	member
Kyoto Animation	Kyoto Animation: Shounen Ken	111.0	1010.0	0.88	1130.0
	Kyoto Animation: Seinen Ken	111.0	1010.0	0.75	1030.0
	Kyoto Animation: Isekai Ken	111.0	1010.0	0.80	1000.0
	Kyoto Animation: Shounen Ken	111.0	1010.0	0.85	1050.0
	Kyoto Animation: Seinen Ken	111.0	1010.0	0.80	1000.0
...
NIKK	NIKK: Seinen Ken	765.0	1.0	1.00	1000.0
	NIKK: Shounen Ken	765.0	1.0	0.00	100.0
	NIKK: Isekai Ken	765.0	1.0	0.00	144.0
	NIKK: Shounen Ken	765.0	1.0	0.00	170.0

Gambar 4.13 Pengelompokan data berdasarkan nama company dan anime
 Gambar 4.13 adalah *source code* pengelompokan data dan bentuk dataframe setelah dikelompokkan berdasarkan nama *company* dan anime yang diproduksi nya. Selanjutnya peneliti akan mengelompokkan data berdasarkan nama *company*, sehingga hanya tersisa data nama *company* yang merupakan data teks/string dan data lainnya yang berupa data angka.

```
dfMean = df.groupby(['companyName', 'animeName']) \
    .mean() \
    .groupby('companyName') \
    .mean().sort_values('favorites', ascending=False)
```

companyName	numberOfAnime	favorites	animeRate	member
Kyoto Animation	131.0	16181.0	7.068015	213817.931298
Madhouse	383.0	13826.0	6.764726	129319.391645
MAPP	75.0	12990.0	6.334133	238802.173333
Wit Studio	60.0	12854.0	6.918735	295230.88235
ufotable	71.0	12156.0	6.803521	203881.492996
...
Warner Bros. Japan	110.0	36.0	6.911780	256121.754237
Sony Music Entertainment	169.0	33.0	6.362485	179987.852071
Production Igd	71.0	33.0	6.889437	16068.507042
4Kids Entertainment	44.0	26.0	6.844091	130996.250000
NIKK	765.0	1.0	2.158392	13666.039216

Gambar 4.14 Pengelompokan data berdasarkan companyName
 Pada gambar 4.14 kita bisa melihat data yang sudah dikelompokkan berdasarkan nama *company*. pada dataframe tersebut semua data angka adalah hasil perhitungan rata-rata dari keseluruhan kolom data. Data tersebut disimpan dalam variabel dfMean. Setelah di proses dataframe memiliki 129 baris dan 4 kolom. Data yang sudah diproses ini selanjutnya akan digunakan untuk proses uji statistik pada tahap selanjutnya.

Uji Korelasi

Data yang telah diproses kemudian diuji dengan metode analisis data uji korelasi pearson, dimana data di kolom favorites pada gambar 4.9 akan menjadi variabel X, sedangkan data pada 4 kolom lainnya, yaitu “numberOfAnime”, “animeRate”, dan “member” menjadi variabel Y Tujuan tes ini adalah untuk mengetahui adakah hubungan korelasi antara popularitas studio animasi dengan anime serta membuktikan nilai koefisien korelasi antara popularitas studio pembuatan animasi dengan anime yang diproduksinya.

Dalam pengujian ini, peneliti mencari nilai Koefisien korelasi yang menunjukkan kuat/tidaknya hubungan linier antara dua variabel. Proses uji korelasi akan dilakukan secara otomatis menggunakan library pandas pada google collaboratory.

```
dfMean=dfMean[["favorites", "numberOfAnime", "animeRate", "member"]]
dfMean.corr()
```

	favorites	numberOfAnime	animeRate	member
favorites	1.000000	0.076949	0.167755	0.273537
numberOfAnime	0.076949	1.000000	-0.071569	0.076327
animeRate	0.167755	-0.071569	1.000000	0.535954
member	0.273537	0.076327	0.535954	1.000000

Gambar 4.15 Proses uji korelasi dengan library pandas

Gambar 4.15 merupakan source code beserta hasil akhir penghitungan koefisien korelasi, untuk melakukan penghitungan koefisien korelasi, peneliti menuliskan variabel “dfMean” untuk menyimpan data hasil penghitungan dan fungsi “.corr()”. dengan menuliskan fungsi “.corr()”, google collaboratory akan langsung menampilkan semua koefisien korelasi data di setiap kolom dengan kolom lainnya secara otomatis dan lengkap.

Pada penelitian ini peneliti berfokus pada hasil koefisien korelasi dari kolom “favorites” yang merupakan jumlah favorite pada studio animasi atau parameter popularitas studio animasi dengan kolom lainnya. Hasil uji korelasi dan interpretasi nilainya dijelaskan pada tabel berikut.

Tabel 4.1 hasil uji korelasi dan interpretasinya

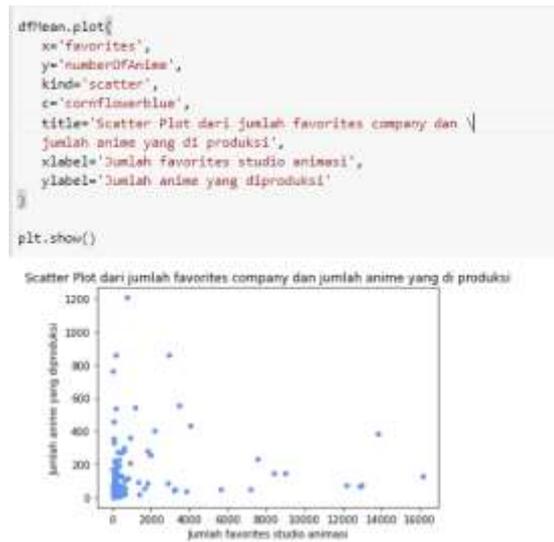
Variabel X	Variabel Y	Nilai koefisien korelasi	Kekuatan korelasi
Jumlah favorites	Jumlah anime	0.076949	Sangat Rendah
Jumlah favorites	Rating anime	0.167755	Sangat Rendah
Jumlah favorites	Jumlah member	0.273537	Rendah

Dari tabel 4.1 diatas, menurut Sugiyono (2013) yang telah dijelaskan sebelumnya pada bab 2.11, dapat disimpulkan bahwa :

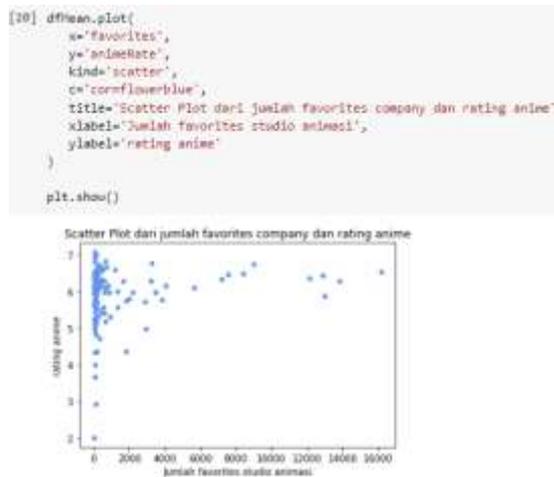
3. Jumlah favorite studio animasi dan jumlah anime memiliki korelasi kuat
4. Jumlah favorite studio animasi dan jumlah rating anime memiliki korelasi sangat rendah
5. Jumlah favorite studio animasi dan jumlah member memiliki korelasi rendah

Visualisasi Data

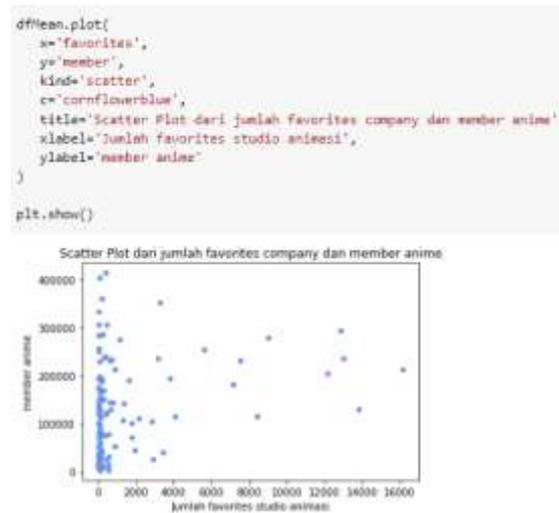
Pada proses ini peneliti akan menampilkan data hasil uji korelasi dalam bentuk scatter plot dengan menggunakan salah satu library python yaitu matplotlib.



Gambar 4.16 Scatter plot dari jumlah favorites company dan jumlah anime
 Gambar 4.16 adalah hasil visualisasi data dalam bentuk scatter plot untuk membandingkan jumlah favorites studio animasi dengan jumlah anime yang diproduksi . disini dapat terlihat ada hubungan linear walaupun beberapa titik masih acak.



Gambar 4.17 Scatter plot dari jumlah favorites company dan rating anime
 Gambar 4.17 adalah hasil visualisasi data dalam bentuk scatter plot untuk membandingkan jumlah favorites studio animasi dengan rating anime yang diproduksi. disini dapat terlihat ada hubungan linear walaupun beberapa titik masih acak.



Gambar 4.18 Scatter plot dari jumlah favorites company dan member anime
 Gambar 4.18 adalah hasil visualisasi data dalam bentuk scatter plot untuk membandingkan jumlah favorites studio animasi dengan member anime yang diproduksi. disini dapat terlihat ada hubungan linear walaupun beberapa titik masih acak.

KESIMPULAN

Berdasarkan hasil analisis hubungan popularitas studio animasi dan anime dengan menggunakan metode pengumpulan data web scraping dan metode analisis data uji korelasi pada situs myanimelist.net. Peneliti mendapatkan kesimpulan sebagai berikut.

1. Metode *web scraping* dapat diimplementasikan pada situs myanimelist.net. peneliti dapat mengumpulkan 17987 baris data dari 3 halaman
2. Hasil uji korelasi menunjukkan bahwa terdapat 3 variabel yang memiliki hubungan korelasi dengan popularitas studio animasi, yaitu jumlah anime, *rating* anime dan *member* dari anime yang diproduksi oleh studio animasi
3. Koefisien korelasi dari hubungan popularitas studio animasi dengan 3 variabel tersebut menunjukkan nilai sebagai berikut:
 - a. Jumlah *favorite* studio animasi dan jumlah anime yang diproduksi: 0.076949
 - b. Jumlah *favorite* studio animasi dan jumlah *rating* anime yang diproduksi: 0.167755
 - c. Jumlah *favorite* studio animasi dan jumlah member dari anime yang diproduksi: 0.273537

Sehingga dapat disimpulkan bahwa popularitas studio animasi memiliki kekuatan hubungan korelasi sangat rendah dengan popularitas anime dan banyaknya anime yang diproduksi.

DAFTAR PUSTAKA

- A. Yani, D. D., Pratiwi, H. S., & Muhandi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem dan Teknologi Informasi*.
- Hajba, G. L. (2018). Using beautiful soup. In *Website Scraping with Python* (pp. 41–96). Apress, Berkeley, CA. https://doi.org/https://doi.org/10.1007/978-1-4842-3925-4_3
- Krotov, V. . & T. M. (2018). Research note: scraping financial data from the web using the R language. *Journal of Emerging Technologies in Accounting*, 15(1), 169-181.
- Purnamasari, Z. (2018). ANALISIS SEMIOTIKA PADA FILM ANIME KIMI NO NA WA . Doctoral Dissertation, PERPUSTAKAAN.
- Python Software Foundation. (2021). beautifulsoup4 . PyPI. Pypi.Org. <https://pypi.org/project/beautifulsoup4/>
- Python Software Foundation. (2022). About PythonTM | Python.org. Python.Org. <https://www.python.org/about/>
- Ramadhan, W. (2021, April 2). Ini Dia Sejarah Anime Hingga Sampai ke Indonesia. *Kreativv*. <https://kreativv.com/sejarah-anime-indonesia/>
- Satria Jati, S., Panuntun, S. B., & Pramana, S. (2021). IMPLEMENTASI WEB SCRAPING DALAM PENGUMPULAN BERITA KRIMINAL PADA MASA PANDEMI COVID-19. *Seminar Nasional Official Statistics*, 2020(1), 300–308. <https://doi.org/10.34123/semnasoffstat.v2020i1.578>
- Schley, M. (2019). Younger Animators Still Struggling Amid Anime Boom. *Japan Times*. <https://www.japantimes.co.jp/culture/2019/05/08/general/>
- Slamet, C., Andrian, R., Maylawati, D. S., Dharmalaksana, W., & Ramdhani, M. A. (2018). Web Scraping and Naïve Bayes Classification for Job Search Engines. *IOP Conference Series: Materials Science and Engineering*, 288. <https://doi.org/10.1088/1757-899X/288/1/012038>
- Sugiyono. (2013). *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D*. Alfabeta.
- Sukardi. (2021). *Metodologi Penelitian Pendidikan: Kompetensi Dan Praktiknya (Edisi Revisi)*. Bumi Aksara.
- Tokyo Otaku Mode. (2015). Exclusive Interview with the Founder of MyAnimeList, a Colossal Site with 120 Million Monthly PVs. *Tokyo Otaku Mode News*. <https://otakumode.com/news/5590fbd763cd06585662ce9d/Exclusive-Interview-with-the-Founder-of-MyAnimeList-a-Colossal-Site-with-120-Million-Monthly-PVs>
- Van Baaren, E. (2021). Python History. *Python Land* . <https://python.land/python-tutorial/python-history>