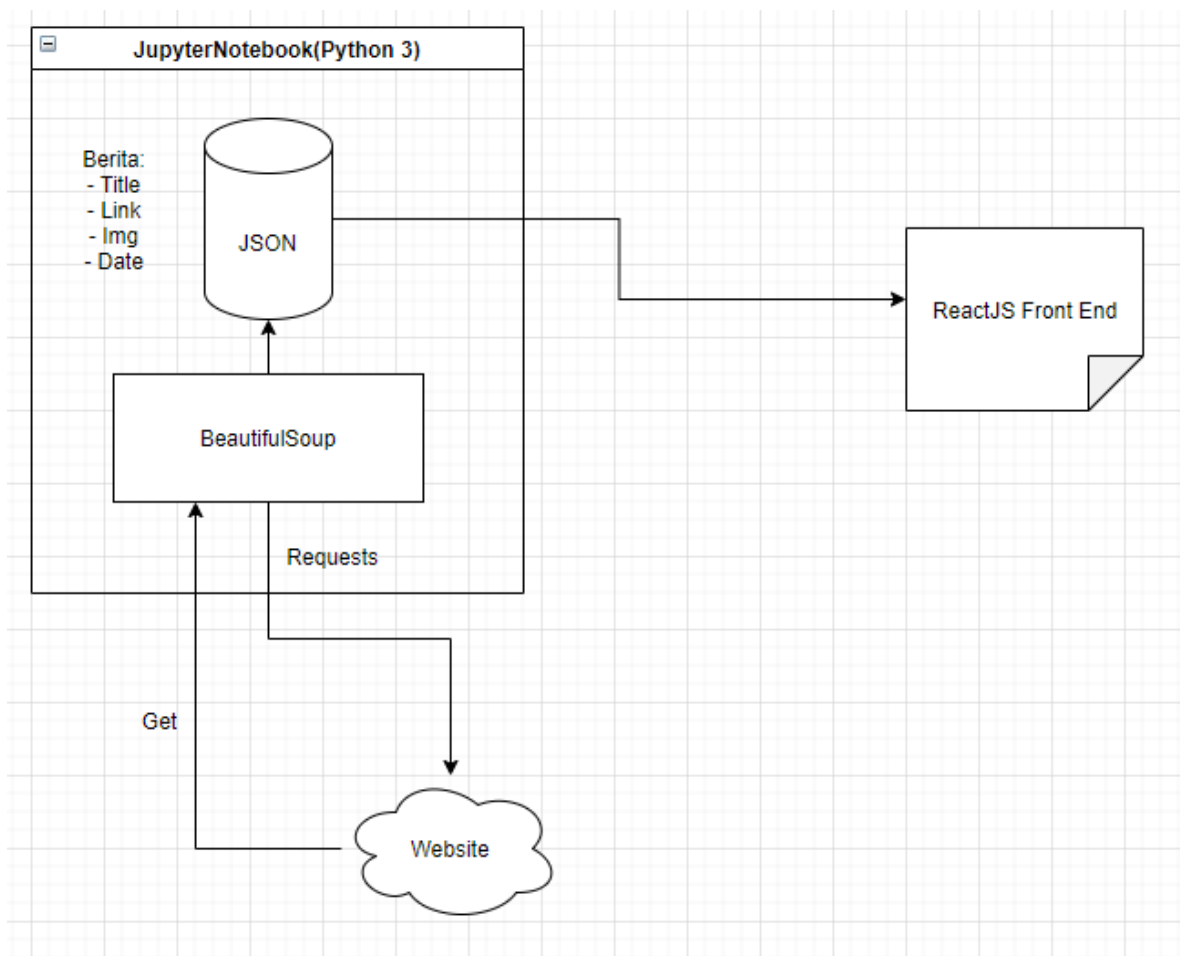


NAMA : RIZALDO ABDULRACHMAN

NIM : 14117152

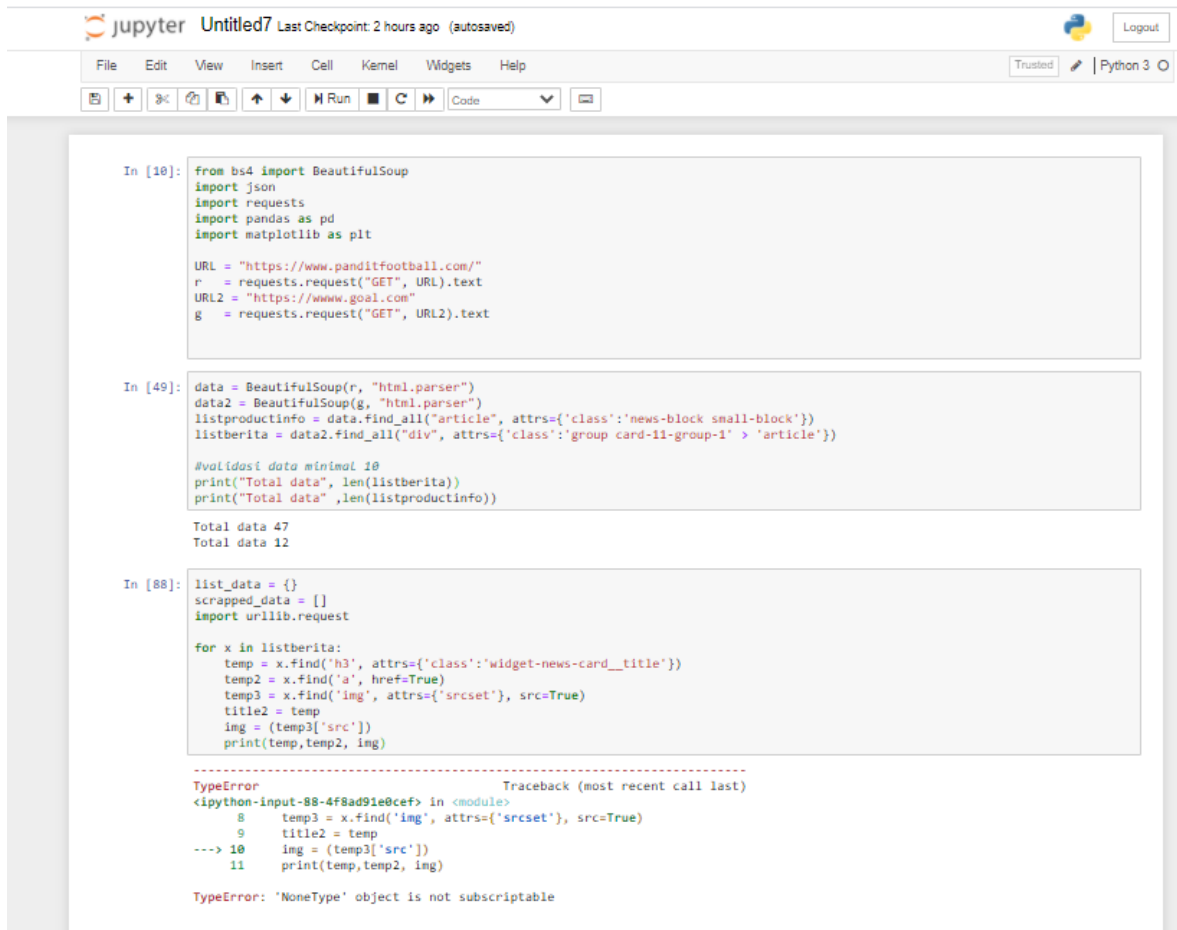
Git : <https://github.com/RizaldoAR/webscrapping-webapp/blob/master/README.md>

WEB SCRAPING APP WITH PYTHON



Gambar 1 : Diagram Web

Untuk scrapping saya menggunakan python 3 di jupyter notebook dengan menggunakan BeautifulSoup.



```
In [10]: from bs4 import BeautifulSoup
import json
import requests
import pandas as pd
import matplotlib as plt

URL = "https://www.panditfootball.com/"
r = requests.request("GET", URL).text
URL2 = "https://www.goal.com"
g = requests.request("GET", URL2).text

In [49]: data = BeautifulSoup(r, "html.parser")
data2 = BeautifulSoup(g, "html.parser")
listproductinfo = data.find_all("article", attrs={'class':'news-block small-block'})
listberita = data2.find_all("div", attrs={'class':'group card-11-group-1' > 'article'})

#validasi data minimal 10
print("Total data", len(listberita))
print("Total data", len(listproductinfo))

Total data 47
Total data 12

In [88]: list_data = {}
scrapped_data = []
import urllib.request

for x in listberita:
    temp = x.find('h3', attrs={'class':'widget-news-card_title'})
    temp2 = x.find('a', href=True)
    temp3 = x.find('img', attrs={'srcset'}, src=True)
    title2 = temp
    img = (temp3['src'])
    print(temp,temp2, img)

-----
TypeError                                Traceback (most recent call last)
<ipython-input-88-4f8ad91e0cef> in <module>
      8     temp3 = x.find('img', attrs={'srcset'}, src=True)
      9     title2 = temp
----> 10     img = (temp3['src'])
      11     print(temp,temp2, img)

TypeError: 'NoneType' object is not subscriptable
```

Gambar 2 : Source code (Python)

```

In [7]: list_data = {}
        scrapped_data = []

        for x in listproductinfo:
            temp = x.find('h3', attrs={'class': 'news-title'})
            temp2 = x.find('a', href=True)
            temp3 = x.find('img', src=True)
            temp4 = x.find('p', attrs={'class': 'simple-share'})
            title = temp.find('a').text
            link = (temp2['href'])
            img = (temp3['src'])
            date = temp4.text
            source = "panditfootball.com"
            title = title.replace('\n', '')
            date = date.replace('\n', '')
            list_data["Title"] = title
            list_data["Link"] = link
            list_data["Img"] = img
            list_data["Date"] = date
            list_data["Source"] = source
            list_datacopy = list_data.copy()
            scrapped_data.append(list_datacopy)

        scrapped_data=json.dumps(scrapped_data)
        print(scrapped_data)

```

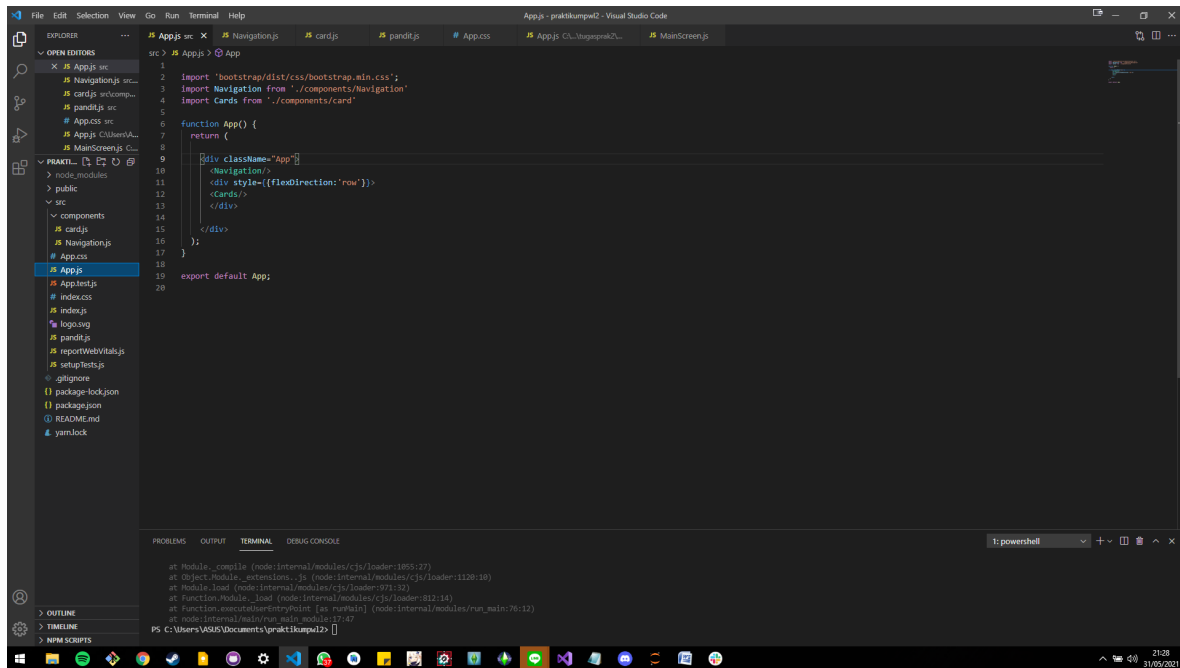
```

[{"Title": "Pekan Terakhir Serie A 2020/21: Siapa Lolos ke Liga Champions?", "Link": "http://www.panditfootball.com/analisa-pertandingan/214390/210523/pekan-terakhir-serie-a-2020-21-siapa-lolos-ke-liga-champions", "Img": "https://img.panditfootball.com/small/Serie%20A.jpg", "Date": "23/05/2021 17:22", "Source": "panditfootball.com"}, {"Title": "Pekan Terakhir Premier League 2020/21: Tiga Klub, Dua Spot UCL, dan Perebutan Golden Boot", "Link": "https://www.panditfootball.com/analisa-pertandingan/214389/210523/pekan-terakhir-premier-league-2020-21-tiga-klub-dua-spot-ucl-dan-perebutan-golden-boot", "Img": "https://img.panditfootball.com/small/PL%20trofi.jpg", "Date": "23/05/2021 17:20", "Source": "panditfootball.com"}, {"Title": "Pekan Terakhir La Liga 2020/21: Siapa Menjadi Juara?", "Link": "https://www.panditfootball.com/analisa-pertandingan/214388/210522/pekan-terakhir-la-liga-2020-21-siapa-menjadi-juara", "Img": "https://img.panditfootball.com/small/PL%20trofi.jpg", "Date": "22/05/2021 21:11", "Source": "panditfootball.com"}, {"Title": "Pemain Spekulatif FPL PanditFootball: Gameweek 37", "Link": "https://www.panditfootball.com/fpl-football-culture/214380/RFH/210518/pemain-spekulatif-fpl-panditfootball-gameweek-37", "Img": "https://img.panditfootball.com/small/WhatsApp%20Image%202021-05-18%20at%2006.10.26%20(1).jpeg", "Date": "18/05/2021 19:34", "Source": "panditfootball.com"}, {"Title": "Penyerang Unggulan FPL PanditFootball: Gameweek 37", "Link": "https://www.panditfootball.com/fpl-football-culture/214379/RFH/210518/penyerang-unggulan-fpl-panditfootball-gameweek-37", "Img": "https://img.panditfootball.com/small/WhatsApp%20Image%202021-05-18%20at%2006.10.26.jpeg", "Date": "18/05/2021 18:18", "Source": "panditfootball.com"}, {"Title": "Gelandang Unggulan FPL PanditFootball: Gameweek 37", "Link": "https://www.panditfootball.com/fpl-football-culture/214378/RFH/210518/gelandang-unggulan-fpl-panditfootball-gameweek-37", "Img": "https://img.panditfootball.com/small/WhatsApp%20Image%202021-05-18%20at%2006.10.24.jpeg", "Date": "18/05/2021 16:14", "Source": "panditfootball.com"}, {"Title": "Kiper dan Bek Unggulan FPL PanditFootball: Gameweek 37", "Link": "https://www.panditfootball.com/fpl-football-culture/214377/RFH/210518/kipper-dan-bek-unggulan-fpl-panditfootball-gameweek-37", "Img": "https://img.panditfootball.com/small/WhatsApp%20Image%202021-05-18%20at%2006.10.25.jpeg", "Date": "18/05/2021 15:47", "Source": "panditfootball.com"}, {"Title": "Emma Hayes, Tokoh Sepakbola Perempuan di Ambang Sejarah", "Link": "https://www.panditfootball.com/cerita/214372/210516/emma-hayes-tokoh-sepakbola-perempuan-di-ambang-sejarah", "Img": "https://img.panditfootball.com/small/Feature-Image-emma.jpg", "Date": "16/05/2021 10:00", "Source": "panditfootball.com"}, {"Title": "Pemain Spekulatif FPL PanditFootball: Gameweek 36", "Link": "https://www.panditfootball.com/fpl-football-culture/214376/RFH/210514/pemain-spekulatif-fpl-panditfootball-gameweek-36", "Img": "https://img.panditfootball.com/small/WhatsApp%20Image%202021-05-14%20at%2014.51.43.jpeg", "Date": "14/05/2021 20:53", "Source": "panditfootball.com"}, {"Title": "Penyerang Unggulan FPL PanditFootball: Gameweek 36", "Link": "https://www.panditfootball.com/fpl-football-culture/214375/RFH/210514/penyerang-unggulan-fpl-panditfootball-gameweek-36", "Img": "https://img.panditfootball.com/small/WhatsApp%20Image%202021-05-14%20at%2014.51.42.jpeg", "Date": "14/05/2021 20:07", "Source": "panditfootball.com"}, {"Title": "Gelandang Unggulan FPL PanditFootball: Gameweek 36", "Link": "https://www.panditfootball.com/fpl-football-culture/214374/RFH/210514/gelandang-unggulan-fpl-panditfootball-gameweek-36", "Img": "https://img.panditfootball.com/small/WhatsApp%20Image%202021-05-14%20at%2014.51.42%20(1).jpeg", "Date": "14/05/2021 18:38", "Source": "panditfootball.com"}, {"Title": "Kiper dan Bek Unggulan FPL PanditFootball: Gameweek 36", "Link": "https://www.panditfootball.com/fpl-football-culture/214373/RFH/210514/kipper-dan-bek-unggulan-fpl-panditfootball-gameweek-36", "Img": "https://img.panditfootball.com/small/WhatsApp%20Image%202021-05-14%20at%2014.51.42.jpeg", "Date": "14/05/2021 17:02", "Source": "panditfootball.com"}]

```

Gambar 3 : Source code (Python)

Setelah dapat outputnya saya pindahkan langsung ke dalam folder master di react, lalu saya baguskan code json nya dengan beautify dari vscode.



Gambar 4 : Source code (App.js)

Untuk menampilkan berita saya menggunakan React + Bootstrap. Untuk menampilkan berita 1 - 1 saya menggunakan wadah card yang dimana ada gambar, title, link, dan tanggal. Pertama saya buat terlebih dahulu components card nya lalu nanti tinggal di panggil di App.js












```

1  import 'bootstrap/dist/css/bootstrap.min.css';
2  import { Navbar, Nav, NavDropdown, Image, Card, Button } from 'react-bootstrap';
3  import Data from './pandit'
4
5
6
7  function Cards() {
8    return (
9
10     <div>
11       <div className="info" style={display: 'flex', flexDirection: 'row'}>
12         <Data.berita.map(info=>
13
14           <Card style={{ width: '18rem', flex: 1 }}>
15             <Card.Img variant="top" src={info.img} />
16             <Card.Body>
17               <Card.Title>{info.Title}</Card.Title>
18               <Card.Text>
19                 {info.Date}
20               </Card.Text>
21               <Button variant="primary" href={info.Link}>Link</Button>
22             </Card.Body>
23           </Card>
24         )
25       </div>
26     </div>
27   );
28 }
29
30 export default Cards;

```

Gambar 5 : Source code (Card.js)

Berikut hasil dari websitenya.

Web Scrapping News Portal										
										
Pekan Terakhir Serie A 2020/21: Siapa Lolos ke Liga Champions?	Pekan Terakhir Premier League 2020/21: Tiga Klub, Dua Spot UCL, dan Perebutan Golden Boot	Pekan Terakhir La Liga 2020/21: Siapa Menjadi Juara?	Pemain Spekulatif FPL PanditFootball all: Gameweek 37	Penyerang Unggulan FPL PanditFootball all: Gameweek 37	Gelandang Unggulan FPL PanditFootball all: Gameweek 37	Kiper dan Bek Unggulan FPL PanditFootball all: Gameweek 37	Emma Hayes, Tokoh Sepakbola Perempuan di Ambang Sejarah	Pemain Spekulatif FPL PanditFootball all: Gameweek 36	Penyerang Unggulan FPL PanditFootball all: Gameweek 36	Gelandang Unggulan FPL PanditFootball all: Gameweek 36
23/05/2021 17:22	23/05/2021 17:20	22/05/2021 21:11	18/05/2021 19:34	18/05/2021 18:18	18/05/2021 16:14	18/05/2021 15:47	16/05/2021 10:00	14/05/2021 20:53	14/05/2021 20:07	14/05/2021 18:38
Link	Link	Link	Link	Link	Link	Link	Link	Link	Link	Link

Gambar 6 :Hasil Website

