JUSIKOM PRIMA (Jurnal Sistem Informasi dan Ilmu Komputer Prima)
Vol. 7 No. 1, August 2023 E-ISSN : 2580-2879

DEVELOPMENT OF STREAMLIT-BASED HIGHER EDUCATIONRANKING INSTRUMENT BOARDS

Glenny Chudra[*1], Alfa Yohannis[2], Richard Setiawan[3]*
[1,2,3]Pradita University
Scientia Business Park, Jl. Gading Serpong Boulevard No. 1, Curug Sangereng, Kec. Klp. Two, Tangerang Regency, SouthTangerang, Indonesia
*E-mail: * Glenny.chudra.s2@student.pradita.ac.id*

ABSTRACT-The dashboard is a tool used to measure and evaluate comparisons of tertiary institutions based on six main aspects, namely teaching, research, industry income, international outlook, and citation scores. (citation score). This research aims to develop an interactive and easy-to-use college ranking dashboard using thePython language. The data obtained will be taken from a website using the scrapping method, and then the data will be combined annually. The method applied in this study uses data life cycle management with essential stages,including collection, processing, analysis, visualization, and data presentation. The results of this study are expected to be able to create a website that can be used by various parties, including prospective students, parents, and researchers who plan to conduct further research. This dashboard lets users obtain relevant information and compare multiple tertiary institutions effectively and efficiently.

Keywords: Dashboard, data analysis, scrapping, college.

INTRODUCTION

Both internationally and nationally, universities have a very high role in developing human resources, scientific progress, and growth for a country[1]. Along with the increase in the number of available tertiary institutions, prospective studentsand parents increasingly need the correct informationin choosing the appropriate tertiary institution[2]

In recent years, web and application technologies have also seen an increase. One of the most increasing is the machine learning and dataanalysis industry which is experiencing rapid growth[3]. One of the things that becomes a tool or technology in the development of data analysis is Streamlit.

Streamlit is a plugin used to create website visualization and data analysis for dashboard purposes using the Python programming.[4]. Streamlit in this study is used to develop a college ranking dashboard that aims to evaluate and compareseveral universities in a way that is easier and easier to access.

In its development, this research also uses life cycle management data so that the data contained in the dashboard can be ascertained and guaranteed the integrity and validity of the existing data.

Through the development of a streamlet-based university ranking instrument board (dashboard), it is hoped that parents and parents of students can be given an overview of university rankings on a global scale, making it easier for researchers to summarize data collection related to tertiary institutions, and higher education stakeholders. to know their positionquickly.

1. Basic Theory

1.1 Pythons

Python is a programming language used to designnlp, data, the Internet of Things, and others[5][6]. Python is an interpreted, high-level, general-purpose programming language created by Guido van Rossumand released in 1991.[7].

1.2 Selenium

Selenium is software that is used for the automatic testing of web applications. One of the libraries in Python is Selenium Python which provides a simple API for writing functional tests using the web Chromedriver[8].

1.3 Scrape

Scrapping or scraping is carried out inretrieving specific data in a semi-structured manner obtained on a website; this website page usually has HTML or XHTML language[9].

In its development, scrapping can transform unstructured data into structured data, namely spreadsheets, CSV, and databases. Scrapping can useany method, and one way is to use the Python programming language[10]

1.4 Research Methodology

Figure 1. Life cycle management data[11]

The research method used in this study is to use data life cycle management (data life cycle management).The steps in this research can be seen in the image below[12]

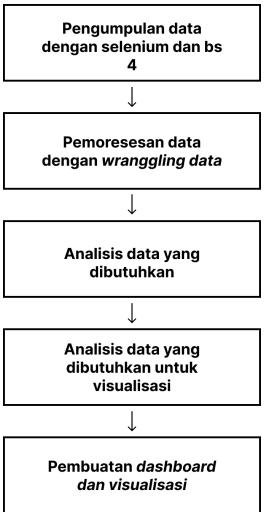

Figure 2. Research method

Using life cycle data, the research method has the following data management steps:

a. *Capture*: the data is collected using the beautifulsoap4 python library and Selenium.
b. *Organize and Store*: the existing data is managed by processing it to make it more tidy, namely by wrangling the data. *Use and Analyze:* Existing data is analyzed, where the required data will be displayed onthe instrument board (dashboard); then, aftercompletion, a Python language program code is created that is used to generate streamlet dashboards.
c. *Shares:* Dashboards created can be shared using custom links provided by the Streamlit library.
d. *Reuse and Maintain*: existing data can then be used for other purposes and develop a dashboard with better features.
e. *Archive and Destroy*: data that is under ten years or a certain period can be deleted so that it can reduce cloud storage.

RESULTS AND DISCUSSION

1.1 Captures

In this study, the data to be collected will use Selenium and Python and will take four values: the research score, citation score, industry income score, and international outlook score. In starting scrape, here we have to provide chromeweb.exe to facilitate capturing data.

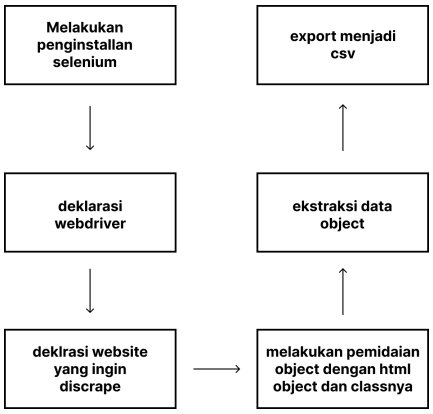

Figure 3. How selenium works

In Figure 3, it is explained that the first time you have to do the installation with Selenium, then Selenium will use the web driver. Therefore, it is important to declare and download it here *web driver. The web driver* will open the claimed website link and scan the data you want to scrape. The observed data will then be extracted and sorted into adata frame to export to CSV. To make it easier to understand, the coding can be seen in Figure 4.

```
from selenium import webdriver
from selenium.webdriver.chrome.service import Service

#deklarasi web driver

webdriver_location = "C:\\Users\\glent\\Downloads\\Selenium-Scrape\\chromedriver.exe"


service = Service(webdriver_location)
driver = webdriver.Chrome(service=service)

#penentuan link yang akan discrape
url_stats = 'https://www.contohtautan.com'

#membuat webdriver chrome dengan nama stats browser dan membuka url diatas dengan webdriver
stats_browser = webdriver.Chrome()
scores_browser.get(url_scores)


#memanggil data dengan object td dan class yang sudah ditentukan
scores_page_html = scores_browser.page_source
scores_page_soup = soup(scores_page_html, 'html.parser')
overall_score_obj = scores_page_soup.findAll("td", {"class":"scores overall-score"})
teaching_score_obj = scores_page_soup.findAll("td", {"class":"scores teaching-score"})
research_score_obj = scores_page_soup.findAll("td", {"class":"scores research-score"})
citations_score_obj = scores_page_soup.findAll("td", {"class":"scores citations-score"})
industry_income_score_obj = scores_page_soup.findAll("td", {"class":"scores industry_income-score"})
international_outlook_score_obj = scores_page_soup.findAll("td", {"class":"scores international_outlook-
score"})
scores_browser.close()

#ekstraksi data

overall_score, teaching_score, research_score, citations_score, industry_income_score,
international_outlook_score = [], [], [], [], [], []
for i in range(len(names_obj)):
    if names_obj[i].a is None:
        continue
    web_address.append('https://www.contohtautan.com' + names_obj[i].a.get('href'))
    rank.append(rank_obj[i].text)
    names.append(names_obj[i].a.text)


    overall_score.append(overall_score_obj[i].text)
    teaching_score.append(teaching_score_obj[i].text)
    research_score.append(research_score_obj[i].text)
    citations_score.append(citations_score_obj[i].text)
    industry_income_score.append(industry_income_score_obj[i].text)
    international_outlook_score.append(international_outlook_score_obj[i].text)

#membuat df dan eksport ke csv
df = pd.DataFrame({
    'overall_score' : overall_score,
    'teaching_score' : teaching_score,
    'research_score' : research_score,
    'citations_score' : citations_score,
    'industry_income_score' : industry_income_score,
    'international_outlook_score' : international_outlook_score,
})

df.to_csv('tahun2013.csv', encoding='utf-16', index=False)
```

Figure 4. Example of scrape coding

1.2 Organize and Store

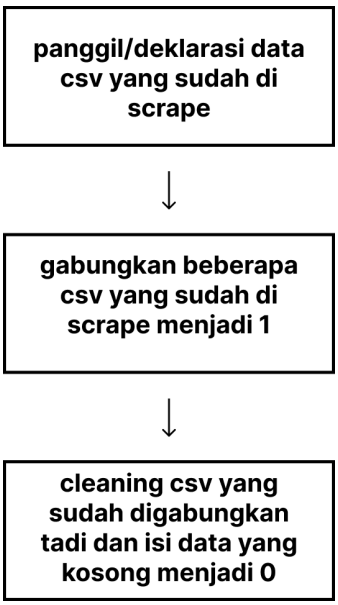

Figure 5. Explanation of organizing and storeprocess

The scraped data will then be managed more neatly. An example of data processing can be seen in Figure 6.

```
import pandas as pd

# Baca data 2014 hingga 2023 csv
df2014 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2014.csv', encoding=file_encoding)
df2015 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2015.csv', encoding=file_encoding)
df2016 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2016.csv', encoding=file_encoding)
df2017 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2017.csv', encoding=file_encoding)
df2018 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2018.csv', encoding=file_encoding)
df2019 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2019.csv', encoding=file_encoding)
df2020 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2020.csv', encoding=file_encoding)
df2021 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2021.csv', encoding=file_encoding)
df2022 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2022.csv', encoding=file_encoding)
df2023 = pd.read_csv(r'C:\Users\Glenny\Documents\GitHub\data2023.csv', encoding=file_encoding)

# Concatenate the dataframes into a single dataframe
merged = pd.concat([df2014, df2015, df2016, df2017, df2018, df2019, df2020, df2021, df2022, df2023])

# Reset the index of the merged dataframe
mergeds = merged.reset_index(drop=True)
merged_df.to_csv('mergeduniversity.csv')
```

Figure 6. Example of organizing data

After all, has been successfully merged using pd. Concat, the next step is to fill in the blank data with the number "0" with the df. Fill (0) command. After it's finished managing, create a directory on the GitHub platform to store CSV data and code that willbe used in the future.

1.3 Use and Analyze

After the data is organized, the data is ready to be used by making an instrument board (dashboard) using Streamlit. Several steps for creating a Streamlit dashboard can be seen in Figure 7.

Figure 7. Streamlight creation flow

After completion, a test was carried out on theinstrument board (*dashboards*) that had been made, namely searching for one of the universities with the keyword "Harvard," which will have an interface thatcan be seen in Figure 9.

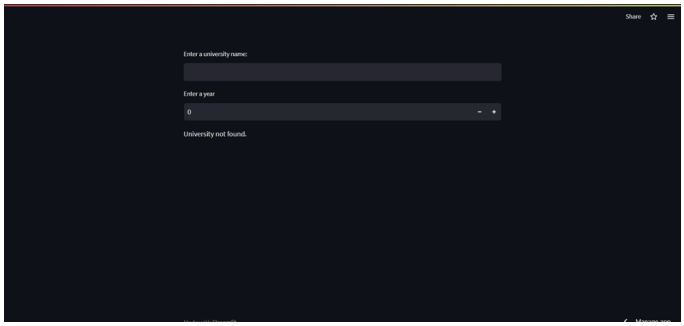

Figure 8. Streamlight dashboard interface

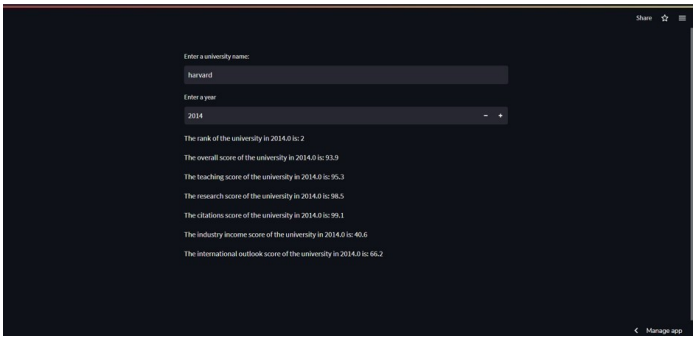

Figure 9. The interface when searching for auniversity

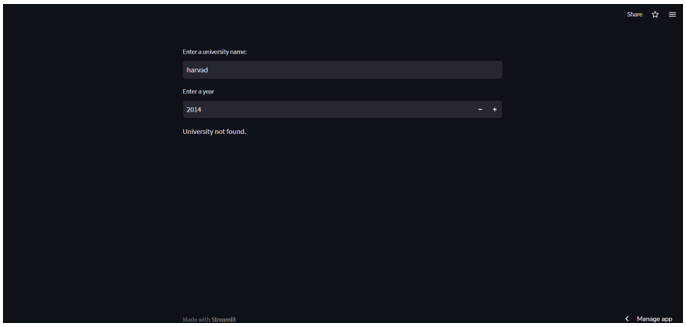

Figure 10. The interface when the university you arelooking for is not found

Figures 8 to 10 are the interface displays for making instrument boards using Streamlit. In Figure 8, it can be seen that the dashboard board still needs to be filled in and is still empty. Figure 9 shows the search results with the Harvard keyword so that it can produce names, rankings, total scores, and 4 determining scores—*rank* university. Figure 10 is the interface where the university and the year you are looking for are not found.

1.4 Shares

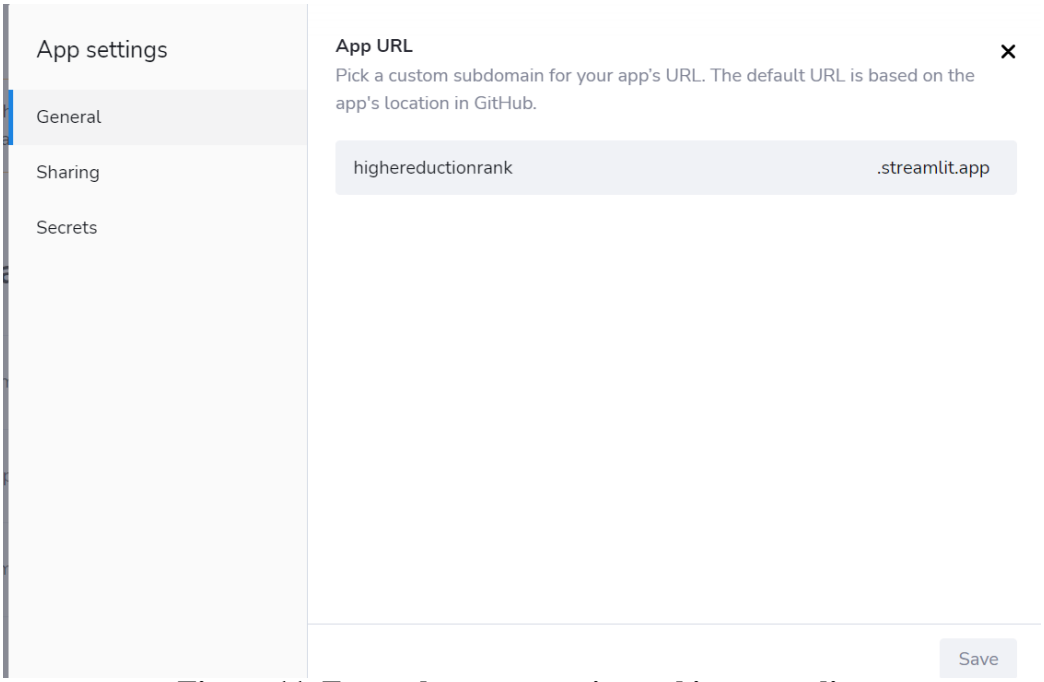

Figure 11. Example*custom settings*url in streamlit

The next stage is to share the results of the instrument boards that we have made. In this stage, streamlet can help us to*custom*appurl so that people who want to use it and are interested can easily see and search for it. An example of how to *custom* URLsis with image 11.

1.5 Reuse and Maintain

At this stage, the existing data can become other data that can be used. An example is to look for a relationship between the year the IT strategy document appeared and university rankings, the relationship between the number of international students and university rankings, and other studies that have something to do with university rankings.

1.6 Archive and Destroy

At this stage, unused data will be archived or destroyed. In the case of university ranking data,whenusing data for only 10 years, namely 2014- 2023, when 2024 is added, 2014 will be archived or destroyed depending on future needs.

1.7 Evaluation of Streamlit

In using Streamlit, of course, there are some advantages and disadvantages. The benefits of Streamlit are that it is easy to use and the existing datais integrated with the GitHub platform, is easy to use,is a library that can be installed with 'pip' from Python, and also a domain that can be modified or custom.

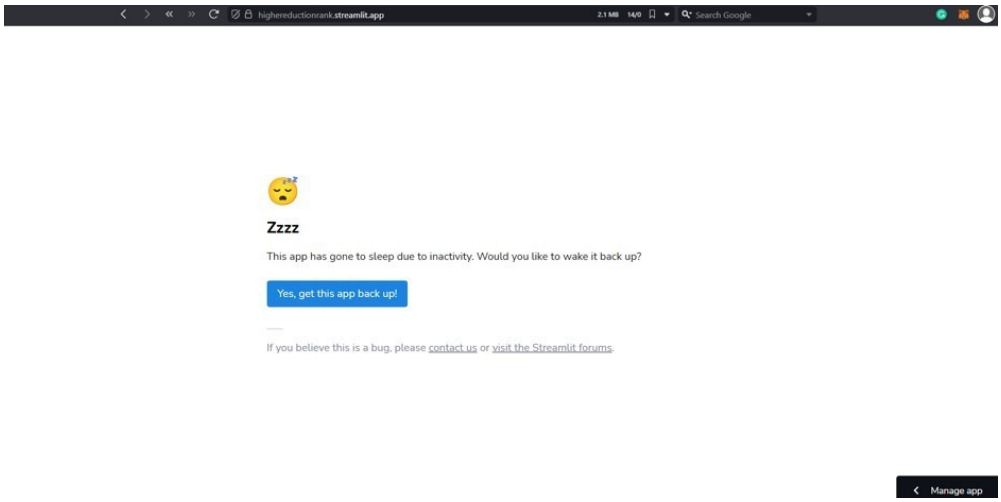

Figure 12. Weaknesses of streamlet

One of the weaknesses of Streamlit is that for 7 consecutive days if it is not used or there are no visitors, it will be immediately deactivated. The Streamlit manager must reactivate it by clicking the backup button, as shown in Figure 12.

CONCLUSION

The conclusion is that the scrapping stage can make it easier to collect data, and Streamlit can makeit easier to use and share dashboards created, but the drawback is that Streamlit has to be opened constantly. If it is not opened, it will be deactivated by Streamlit and must be reactivated manually by clicking the button backup.

In the future, the dashboard can be added withseveral features, such as bar charts or line charts of scores from year to year, comparisons of male and female students with pie charts, and other features that can beautify the dashboard. Moreover, it may be comparable with other platforms used in the production of data dashboards.

BIBLIOGRAPHY

[1] M. Muhardi, "CONTRIBUTION OF EDUCATION IN IMPROVING THE QUALITY OF THE INDONESIAN NATION," vol. XX, no. 4, pp. 478–492, 2005.

[2] RR Putra, "SELECTION OF HIGHER EDUCATION DEPARTMENTS FOR CHILDREN IN SURABAYA," pp. 1–11, 2018.

[3] D. Iskandar, M. Alif Fathoni, and A. Arta Bhrata, "Smart Manufacturing Management System Utilizing Big Data and Machine Learning Algorithms for MSME Production," inform. Mulawarman J.Ilm. Computing Science., vol. 16,no.2,p.96,2021, doi: 10.30872/Jim.v16i2.5258.

[4] Streamlit, “Streamlit Docs,” 2023. https://docs.streamlit.io (accessed Aug. 01, 2023).

[5] Novi Aryani Fitri and Ismaulidia Nurvembrianty, "Virtual Midwife Uses the Polita Chatbot Service Application as a Media for Immunization Information," SATIN - Science and Technology. inf., vol. 7, no. 1, pp. 12–21, 2021, doi: 10.33372/stn.v7i1.678.

[6] R. Muzawi and WJ Kurniawan, "Application of Internet of Things (IoT) in Mobile-Based Light Control Systems," J-SAKTI (Journal of Computer Science and Inform., vol. 2, no. 2, p. 115, 2018, doi: 10.30645/j- sakti.v2i2.75.

[7] H. Halvorsen, Python Programming. 2020.[Online].Available: https://www.halvorsen.blog/documents/programming/python/resources/Python Programming.pdf

[8] B. Muthukadan, “Selenium Python Bindings,” 2014.
[9] AY Syaefulloh, "Web Scraping Using Python," no. March, pp. 1–2, 2020.
[10] R. Mitchell, Web Scraping with Python. 2018. [11]Data.NSW, “Data Management Life Cycle,” 2023. https://data.nsw.gov.au/IDMF/data- management-and-practice/data-management-life-cycle
[11] K. Rahul and RK Banyal, “Data Life Cycle Management in Big Data Analytics,” Procedia Comput. sci., vol. 173, no. 2019, pp. 364–371,2020,doi: 10.1016/j.procs.2020.06.042.