DEVELOPMENT OF STREAMLIT-BASED HIGHER EDUCATION RANKING INSTRUMENT BOARDS

Glenny Chudra*, Alfa Yohannis1, Richard Setiawan3
1,2,3Pradita University
Scientia Business Park, Jl. Gading Serpong Boulevard No. 1, Curug Sangereng, Kec. Klp. Two, Tangerang Regency, South Tangerang, 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 the Python 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 students and parents increasingly need the correct information in 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 data analysis 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 compare several 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 position quickly.
1. Basic Theory

1.1 Python

Python is a programming language used to design NLP, data, the Internet of Things, and others[5][6]. Python is an interpreted, high-level, general-purpose programming language created by Guido van Rossum and 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 Scraping

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

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

1.4 Research Methodology

![Image](image.png)

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

- **Capture**: the data is collected using the beautifulsoup4 python library and Selenium.
- **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 on the instrument board (dashboard); then, after completion, a Python language program code is created that is used to generate streamlet dashboards.
- **Shares**: Dashboards created can be shared using custom links provided by the Streamlit library.
- **Reuse and Maintain**: existing data can then be used for other purposes and develop a dashboard with better features.
- **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.
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. 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 a data frame to export to CSV. To make it easier to understand, the coding can be seen in Figure 4.

![Figure 4. Example of scrape coding](image)

1.2 Organize and Store

```
1. Organize and Store
```

- **Panggil/deklarasi data csv yang sudah di scrape**

  └── gabungkan beberapa csv yang sudah di scrape menjadi 1

  └── cleaning csv yang sudah digabungkan tadi dan isi data yang kosong menjadi 0

![Figure 5. Explanation of organizing and store process](image)
The scraped data will then be managed more neatly. An example of data processing can be seen in Figure 6.

![Figure 6. Example of organizing data](image)

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 will be 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](image)

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

The next stage is to share the results of the instrument boards that we have made. In this stage, streamlet can help us to customize the app URL so that people who want to use it and are interested can easily see and search for it. An example of how to customize URL is 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, when using 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 data is 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.
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 make it 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 with several 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**


