


# Praktikum Pengantar Sains Data STA1381

Mahesa Ahmad Rahmawan | Muhammad Nachnoer Novatron Fitra Arss

# Kontrak Praktikum Sesi UTS



Pertemuan	Topik Responsi	Keterangan
1	Pembagian kelompok dan identifikasi manfaat statistika dan sains data	18 Agustus 2023
2	Data gathering (web scraping) sesuai topik yang dibagikan	25 Agustus 2023
3	Data managing & visualization untuk data yang telah diperoleh	1 September 2023
4	Analisis regresi pada data yang telah diperoleh	8 September 2023
5	Seleksi peubah pada data yang telah diperoleh	15 September 2023
6	Menerapkan metode shrinkage untuk pemodelan pada data yang telah diperoleh	22 September 2023
7	Presentasi hasil kerja kelompok	Infografis
UTS		

- Setiap minggu kelompok membuat laporan hasil yang sudah dikerjakan sesuai topik dalam bentuk power point (convert ke pdf) dan dikumpulkan pada link pengumpulan tugas
- Beberapa kelompok akan dipilih secara acak untuk melakukan presentasi

# Target Responsi 1



➡ Demo Pengolahan (R): Data Gathering (Web Scraping)

➡ Working Group:

Melakukan web scraping pada website: <https://sinta.kemdikbud.go.id/>

Melakukan web scraping data pendukung pada website:  
<https://pddikti.kemdikbud.go.id/>

Target Scraping setiap kelompok ialah: 15 Perguruan Tinggi berstatus PTN-BH

- |  |   |
|--|---|
| 1. Universitas Indonesia (UI).               | 9. Universitas Hasanuddin (Unhas)             |
| 2. Institut Teknologi Bandung (ITB)          | 10. Universitas Sumatera Utara (USU)          |
| 3. Institut Pertanian Bogor (IPB)            | 11. Universitas Pendidikan Indonesia (UPI)    |
| 4. Universitas Padjadjaran (Unpad)           | 12. Universitas Sebelas Maret (UNS) Surakarta |
| 5. Universitas Gadjah Mada (UGM)             | 13. Universitas Andalas (Unand)               |
| 6. Institut Teknologi Sepuluh Nopember (ITS) | 14. Universitas Brawijaya (UB)                |
| 7. Universitas Airlangga (Unair)             | 15. Universitas Negeri Malang (UM).           |
| 8. Universitas Diponegoro (Undip)            |   |



# Responsi #01

Web Scrapping

## Apa itu Web Scrapping?

penggunaan perangkat lunak komputer untuk mengumpulkan data dari seluruh atau beberapa halaman *website* kompetitor

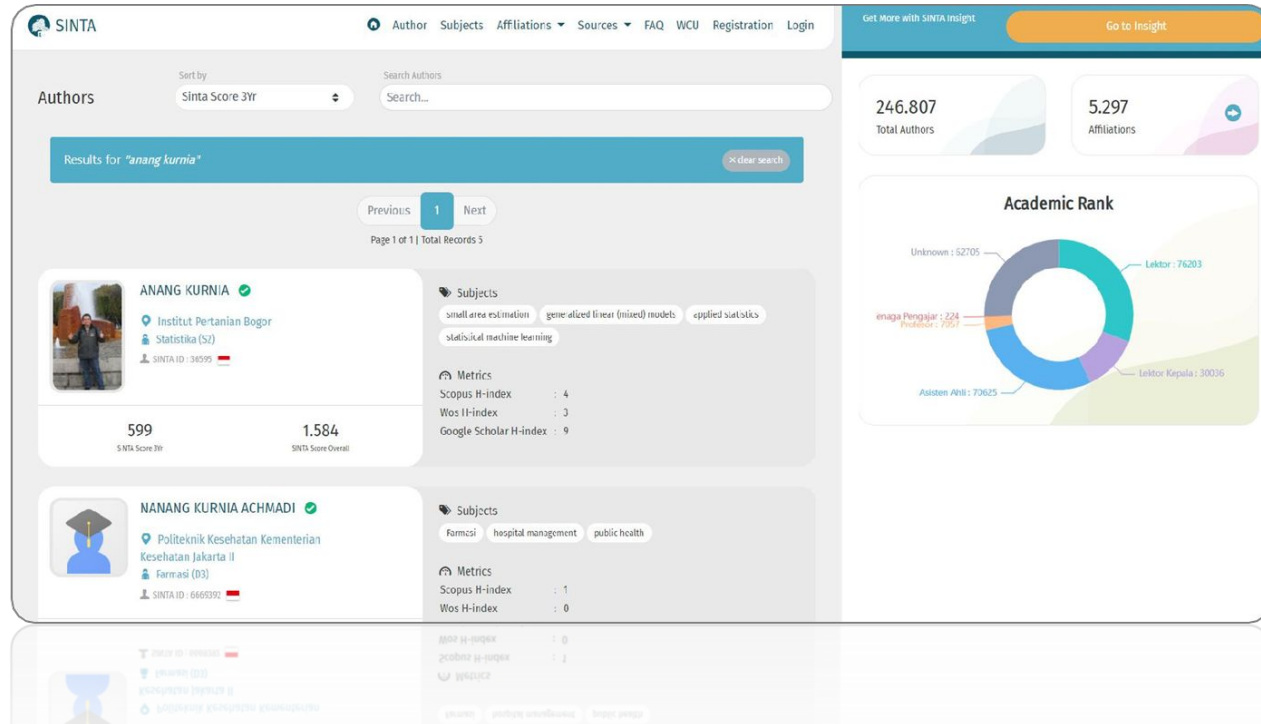


### Salah Satu Teknik Web Scrapping Parsing HTML

mengirimkan request HTTP kepada server yang menyimpan data website tersebut

# Studi Kasus Web Scrapping

Profile Authors dari  
<https://sinta.kemdikbud.go.id>



## Contoh Output

	SINTA_ID	Nama	Universitas	Departemen	SINTA_Score_Overall	SINTA_Score_3Yr	Scopus_Artikel	Scopus_Citation	Scopus_H_Index
1	6059361	WARJITO	Universitas Indonesia	S2 - Teknik Mesin	1.489	843	66	297	9
2	5975420	JOHNY WAHYUADI M	Universitas Indonesia	S2 - Manajemen Integritas Material	2.554	842	91	314	7
3	5975030	DEDE DJUHANA	Universitas Indonesia	S2 - Ilmu Material	2.398	840	76	242	8
4	6772773	JACUB PANDELAKI	Universitas Indonesia	Sp-1 - Radiologi	1.006	838	54	152	6
5	6034474	TEUKU YURI M ZAGLOEL	Universitas Indonesia	S1 - Teknik Industri	1.694	828	76	239	8
6	5986101	MOCHAMMAD CHALID	Universitas Indonesia	S3 - Teknik Metalurgi Dan Material	2.064	825	81	433	9
7	6702176	ROMADHANI ARDI	Universitas Indonesia	S2 - Teknik Industri	1.212	824	57	154	5
8	5986715	MUFTI PETALA PATRIA	Universitas Indonesia	S2 - Ilmu Kelautan	1.320	816	59	142	6
9	6025553	ANTO SULAKSONO	Universitas Indonesia	S3 - Fisika	4.384	816	113	669	15
10	6028142	Agustino Zulys	Universitas Indonesia	S2 - Ilmu Kimia	1.772	815	58	1052	12
11	6059361	WARJITO	Universitas Indonesia	S2 - Teknik Mesin	1.489	843	66	297	9
12	5975420	JOHNY WAHYUADI M	Universitas Indonesia	S2 - Manajemen Integritas Material	2.554	842	91	314	7
13	5975030	DEDE DJUHANA	Universitas Indonesia	S2 - Ilmu Material	2.398	840	76	242	8
14	6772773	JACUB PANDELAKI	Universitas Indonesia	Sp-1 - Radiologi	1.006	838	54	152	6
15	6034474	TEUKU YURI M ZAGLOEL	Universitas Indonesia	S1 - Teknik Industri	1.694	828	76	239	8
16	5986101	MOCHAMMAD CHALID	Universitas Indonesia	S3 - Teknik Metalurgi Dan Material	2.064	825	81	433	9
17	6702176	ROMADHANI ARDI	Universitas Indonesia	S2 - Teknik Industri	1.212	824	57	154	5

# Web Target: [sinta.kemdikbud.go.id](https://sinta.kemdikbud.go.id)

1

Master Authors: <https://sinta.kemdikbud.go.id/authors?page=1>

The screenshot shows the SINTA Authors page. At the top, there's a navigation bar with links like Author, Subjects, Affiliations, Sources, FAQ, WCU, Registration, and Login. Below the navigation bar, there's a search bar and a dropdown menu for 'Sort by Sinta Score 3Yr'. The main content area displays a list of authors. Two authors are visible: YULMIRA YANTI and RAYMOND PRANATA. Each author profile includes a photo, name, affiliation, SINTA ID, and various metrics like Scopus H-index, WoS H-index, and Google Scholar H-index.

Author	SINTA ID	Scopus H-index	WoS H-index	Google Scholar H-index
YULMIRA YANTI	5973268	8.695	10.591	10
RAYMOND PRANATA	6754604	5.569	7.486	29

**Goal:** mendapatkan data authors yang akan dijadikan master

2

Detail Authors:

<https://sinta.kemdikbud.go.id/authors/profile/36595>

The screenshot shows the SINTA Author profile page for ANANG KURNIA. The page includes a profile picture, name, affiliation (Institut Pertanian Bogor), and SINTA ID (36595). It also displays various metrics and a summary of research output. The 'Summary' section includes a donut chart for Article Quartile, a line chart for Latest number of publications, and a table for Scopus, GScholar, and WOS metrics.

Metric	Scopus	GScholar	WOS
Article	54	170	26
Citation	67	372	37
Cited Document	29	86	12
H-index	4	9	3
i10-index	1	9	1
G-index	1	1	1

**Goal:** SINTA ID yang diperoleh dari Master Authors akan digunakan untuk scrap data authors yang lebih detail



Menggunakan



`library(tydiverse)`



`library(rvest)`

rvest helps you scrape (or harvest) data from web pages. It is designed to work with [magrittr](#) to make it easy to express common web scraping tasks, inspired by libraries like [beautiful soup](#) and [RoboBrowser](#).



# 1

## Master Authors

- 1.) Untuk memperoleh list authors (pada kasus ini) dapat diperoleh dengan melakukan scrapping pada url = <https://sinta.kemdikbud.go.id/authors>.
- 2.) Page ini berisikan profil authors yang tidak detail dan bukan merupakan data tabular setiap page berisi 10 authors
- 3.) Untuk memperoleh authors selanjutnya, disediakan tombol pagination sehingga perlu dilakukan scrapping pada setiap page menggunakan fungsi pengulangan

- 1.) Detail dari author dapat dilihat pada page 'https://sinta.kemdikbud.go.id/authors/profile/5973268'. Number pada akhir url merupakan Profile ID atau SINTA ID. untuk mengetahui mengetahui SINTA ID dari authors maka pada kasus ini menggunakan data master authors yang telah dilakukan scrapping sebelumnya.
- 2.) secara umum proses scrapping pada Authors - Detail sama dengan Master Authors. Adapun perbedaannya, pada Authors Detail tidak terdapat pagination sehingga scrapping cukup dilakukan dengan menggunakan pengulangan page berdasar Profile ID yang disematkan pada akhir url

# Pararel 1

A	B	C	D	E
No	NIM	Nama	Status	Kelompok
4	G1401211011	MUHAMMAD	Mayor	1
23	G1401211058	Alfikri Ihsan	Mayor	1
2	G1401201021	Nur Hakim	Mayor	1
40	G1401211099	Muhammad Ha	Mayor	1
41	G1401211100	RANI YASMIN	Mayor	1
36	G1401211088	Amalia Safira V	Mayor	1
35	G1401211087	Pingkan Febbe	Mayor	1
6	G1401211019	Natasha Muti F	Mayor	1
21	G1401211054	ZAFIRA ILMU	Mayor	1
37	G1401211089	Hilmi Awfa Ab	Mayor	1
28	G1401211068	Doni Oktavian	Mayor	1

13	14	G1401211042	INDRA MAUI	Mayor	2
14	3	G1401201038	Irpando Sagala	Mayor	2
15	7	G1401211027	Adisti Suci Ra	Mayor	2
16	27	G1401211067	RIFQI RUSTU	Mayor	2
17	38	G1401211094	Hanifa Rahma	Mayor	2
18	30	G1401211077	Fadly Mocham	Mayor	2
19	1	G1401201012	Tiara Ayu Perti	Mayor	2
20	44	F34188006	Muhammad Fa	SC	2
21	25	G1401211062	Muhammad Fa	Mayor	2
22	18	G1401211049	FEBRI DWI C	Mayor	2
23	19	G1401211051	Muhamad Faw	Mayor	2



24	17	G1401211048	Rafee Aziz Pra	Mayor	3
25	10	G1401211033	NAZUWA AU	Mayor	3
26	13	G1401211040	Raziqizzan Put	Mayor	3
27	42	G1401211104	Jonathan Hizki	Mayor	3
28	11	G1401211034	Dindana Fitriar	Mayor	3
29	34	G1401211086	Ubaidillah Al F	Mayor	3
30	16	G1401211046	AZIZAH AMA	Mayor	3
31	8	G1401211030	Rheyhan Fahry	Mayor	3
32	32	G1401211082	Windi Gunawa	Mayor	3
33	5	G1401211013	REYND HAM	Mayor	3
34	12	G1401211039	HAKIM ZOEL	Mayor	3

35	29	G1401211070	Kaylila Kireina	Mayor	4
36	33	G1401211083	Septiranny Riz	Mayor	4
37	24	G1401211060	Adinda Putri A	Mayor	4
38	15	G1401211045	Azzahra Adelia	Mayor	4
39	20	G1401211052	RAVI MAHES	Mayor	4
40	31	G1401211081	PRATAMA FA	Mayor	4
41	26	G1401211064	JONATHAN M	Mayor	4
42	39	G1401211098	Fajryanti Kusu	Mayor	4
43	43	G1401211106	Megawati Roit	Mayor	4
44	9	G1401211032	Butsainah Taqi	Mayor	4
45	22	G1401211056	NASWA NABI	Mayor	4

## Pararel 2

1	No	NIM	Nama	Status	Kelompok
2	45	G1401211107	Yasmin Azimal	Mayor	1
3	43	G1401211095	RAFI AKBAR	Mayor	1
4	24	G1401211028	Nadila Putri Fa	Mayor	1
5	22	G1401211025	FEDORA ILAL	Mayor	1
6	8	G1401211008	Nabil Naufal	Mayor	1
7	1	G1401211001	Karimatu Ain	Mayor	1
8	3	G1401211003	LUTFI SYAHF	Mayor	1
9	2	G1401211002	Diva Nisfu Mu	Mayor	1
10	20	G1401211023	ARFIAH KAN	Mayor	1
11	11	G1401211012	SYIFA KHAIR	Mayor	1
12	10	G1401211010	MUTIARA AN	Mayor	1

13	17	G1401211020	UTWANG NUI	Mayor	2
14	32	G1401211043	YOGI NUR HA	Mayor	2
15	29	G1401211037	Zulfa Hafizhoh	Mayor	2
16	12	G1401211014	GLADYS ADY	Mayor	2
17	37	G1401211066	OKTAVIA GA	Mayor	2
18	30	G1401211038	DHIYA KHAL	Mayor	2
19	18	G1401211021	RADHITYA H	Mayor	2
20	33	G1401211044	Rafli Radithya	Mayor	2
21	26	G1401211031	Muhammad Lu	Mayor	2
22	5	G1401211005	Cahyani Dyah	Mayor	2
23	36	G1401211065	Vita Rizkyana	Mayor	2





24	6	G1401211006	ANGGA FATE	Mayor	3
25	31	G1401211041	Andi Fatihatul	Mayor	3
26	14	G1401211016	Aida Darajati	Mayor	3
27	21	G1401211024	DAVINA RAC	Mayor	3
28	23	G1401211026	Salsabila Dwi I	Mayor	3
29	38	G1401211072	IGNACIA MA	Mayor	3
30	4	G1401211004	Asfiah Adiba	Mayor	3
31	27	G1401211035	Dinda Khamila	Mayor	3
32	44	G1401211105	Angel Martha I	Mayor	3
33	16	G1401211018	Muhamad Farr	Mayor	3
34	13	G1401211015	FARIK FIRSTI	Mayor	3

35	35	G1401211063	ALFIAH AYU	Mayor	4
36	42	G1401211084	ANIS SULIST	Mayor	4
37	9	G1401211009	Muhammad Ri	Mayor	4
38	7	G1401211007	Farhan Abdilla	Mayor	4
39	28	G1401211036	AKMAL BASI	Mayor	4
40	34	G1401211057	FARREL GILE	Mayor	4
41	39	G1401211073	Kamilah Nurul	Mayor	4
42	25	G1401211029	KHENI HIKM	Mayor	4
43	41	G1401211076	Aisyah Nuruzz	Mayor	4
44	19	G1401211022	Much.Fazrin S	Mayor	4
45	15	G1401211017	Dewi Kunthi S	Mayor	4
46	40	G1401211074	Reyzha Siva D	Mayor	4



Sekian Terima Kasih