

Liga Indonesia

Web Scraping by Billy, G1501231034



Content

What you'll learn about today

- Project Web Scarping
- Apa Itu Liga Indonesia
- Content web scraping
- Gambaran Umum Hasil Web Scraping
- Lampiran Sintaks

Project Web-Scraping

Web Scraping :

[https://ligaindonesiabaru.com/table/index/BRI_LIGA_1_2022-2023.](https://ligaindonesiabaru.com/table/index/BRI_LIGA_1_2022-2023)

website tersebut merupakan website "Liga Indonesia" yang menyediakan informasi mengenai berita terkini seputar Liga Indoneisa, Jadwal, Klasemen, dan lain-lain.



content web-scraping



Menang : Frekuensi menang dari sebuah tim
Seri : Frekuensi seri dari sebuah tim
Kalah : Frekuensi kalah dari sebuah tim
GF : Goals For, frekuensi gol yang berhasil dicetak tim
GA : Goals Against. frekuensi total kebobolan gol oleh pihak lawan.
GD : Goals Difference, selisih antara GF dan GA
Poin : Jumlah poin yang telah dikumpulkan oleh setiap tim sepanjang musim

HASIL WEB SCRAPING

dengan R studio

team <chr>	menang <dbl>	kalah <dbl>	seri <dbl>	GF <dbl>	GA <dbl>	GD <dbl>	poin <dbl>
PSM MAKASSAR	22	3	9	63	28	35	75
PERSIJA JAKARTA	20	8	6	47	27	20	66
PERSIB BANDUNG	19	10	5	54	50	4	62
BORNEO FC SAMARINDA	16	9	9	64	40	24	57
BALI UNITED FC	16	12	6	67	53	14	54
PERSEBAYA	15	12	7	52	45	7	52
BHAYANGKARA PRESISI INDONESIA FC	15	13	6	53	44	9	51
MADURA UNITED FC	14	11	9	39	36	3	51
PERSITA	13	13	8	43	46	-3	47
PERSIS SOLO	11	12	11	50	47	3	44
PERSIK KEDIRI	12	14	8	42	43	-1	44
AREMA FC	12	16	6	32	40	-8	42
PSIS SEMARANG	12	17	5	44	53	-9	41
PERSIKABO 1973	11	15	8	43	48	-5	41
PS BARITO PUTERA	10	16	8	44	55	-11	38

1-15 of 18 rows

Previous [1](#) [2](#) Next

HASIL WEB SCRAPING

dengan mongodb

The screenshot shows the MongoDB Compass interface. The left sidebar lists databases: MDS database, admin, config, local, sample_mflix, sepakbola (which is selected and highlighted in green), and test. The main area shows the 'sepakbola' database, then the 'liga' collection, which is CLUSTERED. The 'Documents' tab is selected, showing 11 documents. A search bar at the top says 'Type a query: { field: 'value' } or Generate query'. Below the search bar are buttons for ADD DATA, EXPORT DATA, UPDATE, and DELETE. The results section displays two documents:

```
_id: ObjectId('6666e9c13940de3fe0072061')
team : "PERSEBAYA"
menang : 15
kalah : 12
seri : 7
GF : 52
GA : 45
GD : 7
poin : 52

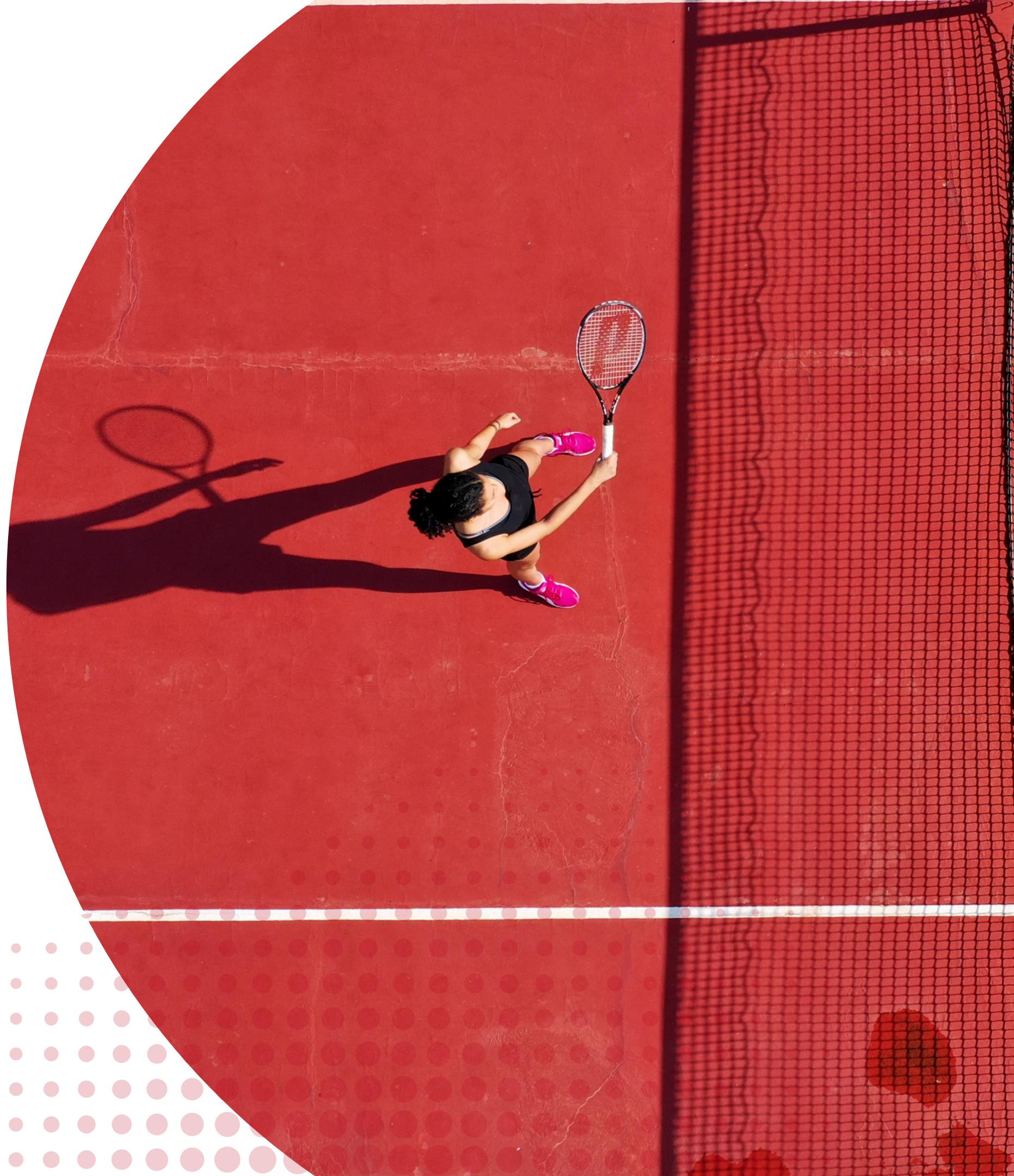
_id: ObjectId('6667af7ef8b2d1ac79086de1')
team : ""
BORNEO FC SAMARINDA"
menang : 16
kalah : 9
seri : 9
GF : 64
GA : 40
GD : 24
poin : 57
```

At the bottom, a dark bar indicates the connection status: >_MONGOSH.

How Our Sponsors Benefit

Highly visible marketing
on a cheerful event for
positive brand feelings

Present with ease and wow any audience with
Canva Presentations. Choose from over a
thousand templates to fit any objective or topic.





Fresh social media **content** and engagement

Present with ease and wow any audience with Canva Presentations. Choose from over a thousand templates to fit any objective or topic.

Hundreds of **customer leads** for each sponsor

Present with ease and wow any audience with Canva Presentations. Choose from over a thousand templates to fit any objective or topic.

Thank you!

Billy, G1501231034

Lampiran

```
message('Loading Packages')
library(rvest)
library(tidyverse)
library(mongolite)
library(httr)
message('Scraping Data')
url_liga <-"https://ligaindonesiabaru.com/table/index/BRI_LIGA_1_2022-2023"
page_liga<-read_html(url_liga)
team <- page_liga %>%
  html_elements(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "noborder", " " ))]//a') %>%
  html_text() %>%
  gsub("\n", "", .)
team<-as.vector(team)
team
menang <- page_liga %>%
  html_elements(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "text-center", " " ))]//*[contains(concat( " ", @class, " " ), concat( " ", "noborder", " " )) and (((count(preceding-sibling::*) + 1) = 4) and parent::*)]') %>%
  html_text() %>%
  gsub("\n", "", .)
menang<-(as.vector(as.numeric(menang)))
menang
```

Lampiran

```
seri<-page_liga %>%
  html_elements(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "text-center", " " ))]//*[contains(concat( " ", @class, " " ), concat( " ", "noborder", " " )) and (((count(preceding-sibling::*) + 1) = 5) and parent::*)])' %>%
  html_text() %>%
  gsub("\n", "", .)
seri<-as.vector(as.numeric(seri))
seri

kalah<-page_liga %>%
  html_elements(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "text-center", " " ))]//*[contains(concat( " ", @class, " " ), concat( " ", "noborder", " " )) and (((count(preceding-sibling::*) + 1) = 6) and parent::*)])' %>%
  html_text() %>%
  gsub("\n", "", .)
kalah<-as.vector(as.numeric(kalah))

GF<-page_liga %>%
  html_elements(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "text-center", " " ))]//*[contains(concat( " ", @class, " " ), concat( " ", "noborder", " " )) and (((count(preceding-sibling::*) + 1) = 7) and parent::*)])' %>%
  html_text() %>%
  gsub("\n", "", .)
GF<-as.vector(as.numeric(GF))
```

Lampiran

```
GA<-page_liga %>%
  html_elements(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "text-center", " " ))]//*[contains(concat( " ", @class, " " ), concat( " ", "noborder", " " )) and (((count(preceding-sibling::*) + 1) = 8) and (parent::*))]' %>%
  html_text() %>%
  gsub("\n", "", .)
GA<-as.vector(as.numeric(GA))

GD<-page_liga %>%
  html_elements(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "text-center", " " ))]//*[contains(concat( " ", @class, " " ), concat( " ", "noborder", " " )) and (((count(preceding-sibling::*) + 1) = 9) and (parent::*))]' %>%
  html_text() %>%
  gsub("\n", "", .)
GD<-as.vector(as.numeric(GD))

poin<-page_liga %>%
  html_elements(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "text-center", " " ))]//*[contains(concat( " ", @class, " " ), concat( " ", "noborder", " " )) and (((count(preceding-sibling::*) + 1) = 10) and (parent::*))]' %>%
  html_text() %>%
  gsub("\n", "", .)
poin<-as.vector(as.numeric(poin))
```

Lampiran

```
liga1<-data.frame(  
  team,  
  menang,  
  kalah,  
  seri,  
  GF,  
  GA,  
  GD,  
  poin  
)  
liga1  
  
pilih <- sample(1:18, 1, replace=FALSE)  
liga1 <- liga1[pilih,]  
  
message('Input Data to MongoDB Atlas')  
atlas_conn <- mongo(  
  collection = Sys.getenv("ATLAS_COLLECTION"),  
  db       = Sys.getenv("ATLAS_DB"),  
  url     = Sys.getenv("ATLAS_URL")  
)  
  
atlas_conn$insert(liga1)  
rm(atlas_conn)
```