

Data Scientist dan R



Halo, selamat datang di kelas pengenalan DQLab yang menandai awal perjalanan kamu sebagai seorang **Data Scientist**!

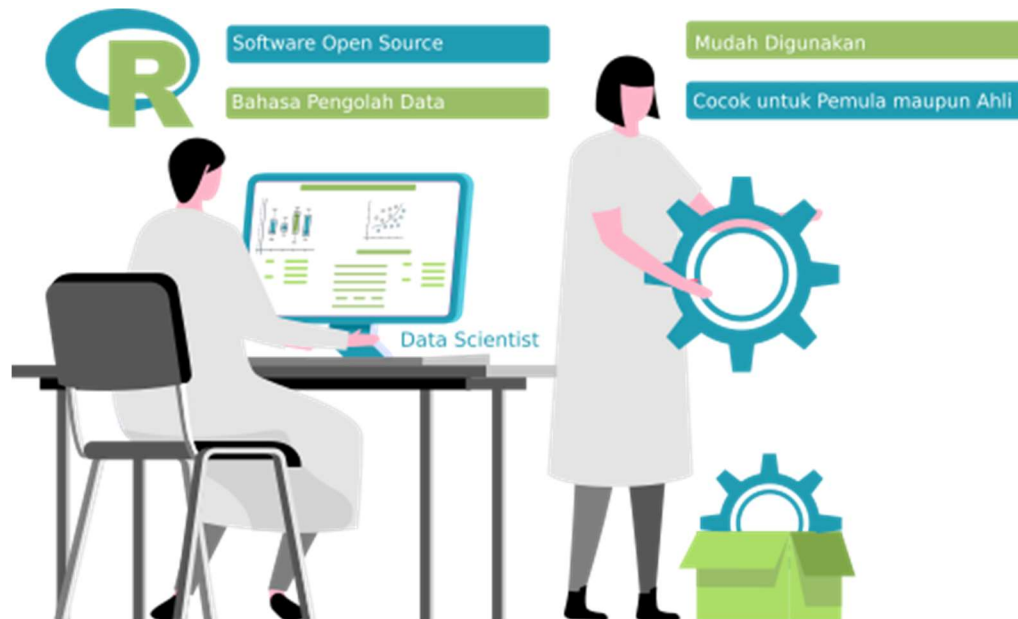
Data Scientist merupakan profesi yang sangat menjanjikan saat ini. Sebagai seorang data scientist tentunya kamu akan bekerja dengan cara merubah data-data yang sebelumnya tidak kelihatan berguna menjadi informasi yang berharga.

Untuk mewujudkan hal itu, seorang data scientist memerlukan *tool* (kakas atau peralatan) dalam bentuk sebuah perangkat lunak (*software*). **Software** yang diperlukan diusahakan cukup simpel namun kuat (*simple but powerful*) dalam penggunaannya.

Nah, *software* yang dimaksud itu dinamakan dengan R. R merupakan *software* pemrograman untuk pengolahan dan pemodelan berbasis data.

Mari kita pelajari lebih dalam mengenai R di dalam modul pembelajaran ini.

Apa dan kenapa R?

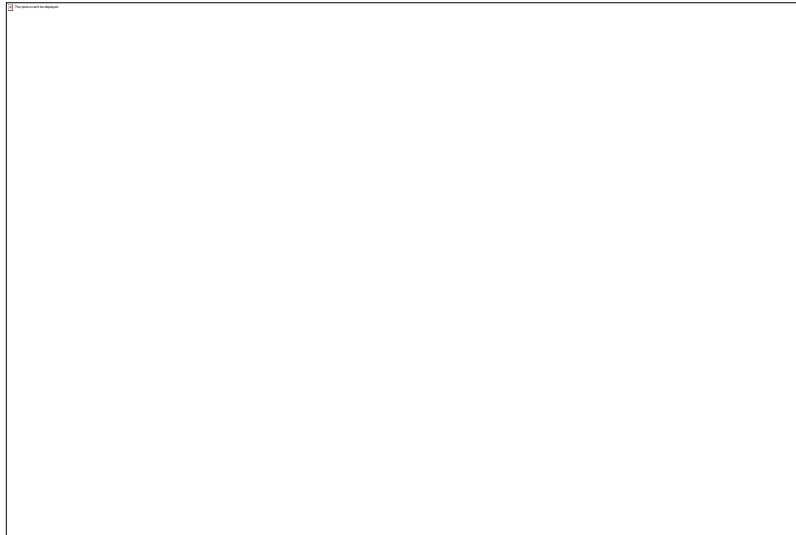


R adalah nama sebuah bahasa pemrograman sekaligus *software* untuk pengolahan data dan grafik.

R sangat populer saat ini karena tiga hal berikut:

- Banyak pilihan pengolahan data dengan jumlah fitur yang sangat **komplit** - dari grafik sampai machine learning.
- **Lebih cepat** dipelajari dan dijalankan untuk mengolah data dibandingkan dengan bahasa lain.
- R bersifat **gratis** dan *open source* yang artinya, tidak perlu biaya lisensi yang biasanya sangat mahal untuk software pengolahan data.

Perusahaan mana yang sudah menggunakan R?



“Tidak ada asap tanpa api”, demikian suatu peribahasa yang berarti tidak ada rumor tanpa ada fakta yang menyertainya.

Kepopuleran bahasa R disebabkan karena penggunaannya pada berbagai perusahaan besar dunia.

Perusahaan besar yang dimaksud diantaranya adalah AirBnB untuk *data science*, Microsoft untuk menambahkan fungsionalitas di produk-produknya, Uber untuk analisa statistik, Facebook untuk behavior analysis, dan lain-lain.

“Hello World” di R

Dalam setiap pembelajaran pemrograman, menampilkan tulisan “Hello World” sebagai tradisi dunia pada saat mulai belajar pemrograman. :)

Mari kita lakukan hal yang sama dengan menggunakan bahasa **R** di laman ini melalui **Code Editor**.

Tugas Praktek

Ketikkanlah “**Hello World**” (diketik persis sama dengan yang dicontohkan baik huruf besar, huruf kecil dan tanda kutip yang dimilikinya) pada bagian **Code Editor**.

Klik tombol **Run** (Angka 1) untuk menjalankan R dan mengeluarkan hasil di bagian **Console**.

Untuk memeriksa hasil yang telah kamu kerjakan apakah sudah benar dan sekaligus agar dapat melanjutkan ke bagian berikutnya, kliklah tombol **Submit** (Angka 2).



```
"Hello World"
```

```
## [1] "Hello World"
```

Praktek: Perhitungan Sederhana

Mari kita lanjutkan kembali praktek sederhana dengan bahasa R. Saat ini kamu diminta untuk melakukan perhitungan sederhana yang berupa penjumlahan sederhana.

Tugas Praktek

Ketikkan pada code editor penjumlah dua bilangan bulat yaitu: **10 + 7**, dan jalankan dengan tombol **Run**.

Jika lancar kamu menjalankannya, hasilnya akan terlihat pada **Console** sebagai berikut.

```
> 10+7  
[1] 17
```

Baris kedua yaitu 17 merupakan hasil perhitungan dari penjumlahan dua bilangan 10 + 7. Abaikan simbol > dan

1

untuk saat ini.

Klik tombol **Submit** untuk melanjutkan ke subbab selanjutnya.

```
10+7
```

```
## [1] 17
```

Variabel dalam R

Penggunaan variabel pada pemrograman sangat penting untuk menampung angka maupun teks di R dengan suatu **nama**.

Tugas Praktek

Pada **Code Editor** ketikkan variabel **a** dengan nilai sebagai berikut:

```
a = 5
```

Ini artinya variabel **a** diisi dengan angka 5. Selanjutnya, tambahkan lagi di baris berikutnya pada Code Editor perintah berikut ini

```
print(a)
```

Ini digunakan untuk mencetak isi variabel **a**.

Klik tombol **Run** dan amati yang dimunculkan di console.

Terkakhir, klik **Submit Code** untuk pengecekan dan melanjutkan ke bagian berikutnya.

```
a=5  
print(a)  
## [1] 5
```

Comment pada R

Comment merupakan teks untuk menambahkan keterangan pada kode yang ditulis. Penambahan keterangan ini akan mengingatkan kita kembali ketika membuka code tersebut di suatu waktu di masa depan.

Comment tidak dianggap sebagai code yang dapat dieksekusi.

Penggunaan *comment* dalam bahasa R dapat dilakukan dengan mengawali suatu teks dengan tanda '#'.

Berikut ini merupakan contoh penggunaan *comment*

```
10 + 7 #Ini adalah baris komentar
```

atau seperti ini

```
#Ini adalah baris komentar.  
10 + 7
```

Praktek: Penggunaan Comment

Ketikkanlah komentar setelah perhitungan matematika seperti yang dicontohkan berikut ini pada Code Editor.

```
10 + 7 #Ini adalah baris komentar
```

Cobalah dijalankan dengan menekan tombol Run, dan jika lancar, maka kamu akan mendapatkan hasil berikut

```
> 10 + 7 #Ini adalah baris komentar  
[1] 17
```

Dari proses ini terlihat bahwa *comment* tidak diproses oleh R, jadi yang ditampilkan hanya hasil perhitungan penjumlahan matematikanya saja.

```
10+7 #ini adalah baris komentar
```

```
## [1] 17
```

Vector pada R

Vector merupakan sebuah struktur data yang dapat menyimpan lebih dari satu data yang akan digunakan di R.

Penggunaannya sangat sederhana, yaitu menggunakan fungsi **c** disertai data-data yang ingin disimpan. Perhatikan potongan kode berikut

```
c(5, 10, 20)
```

ini artinya kamu akan menyimpan nilai 5, 10 dan 20 dalam satu struktur yang disebut vektor. Selanjutnya, jika kamu ingin menyimpan rangkaian angka yang terurut, misalkan angka 1 sampai dengan 20, maka kamu dapat mengetikkan potongan kode berikut

```
c(1:20)
```

Praktek: Penggunaan Vector - Bagian Satu

Ketikkanlah perintah `c(3, 10, 15)` pada Code Editor

```
c(3, 10, 15)
```

Klik tombol **Run** dan perhatikan *output* yang dihasilkan pada bagian **Console**. Di sini dapat kamu melihat hasil sebagai berikut

```
> c(3, 10, 15)
[1] 3 10 15
```

Terlihat bahwa perintah `c(3, 10, 15)` ini membuat tiga urutan angka yaitu 3, 10 dan 15 yang disimpan dan ditampilkan secara bersamaan dalam suatu vector.

```
c(3,10,15)
```

```
## [1] 3 10 15
```

Praktek: Penggunaan Vector - Bagian Dua

Selain mengetikkan satu per satu data pada vector, kita juga dapat membuat urutan data dengan menggunakan operator titik dua. Cobalah ketikkan perintah berikut pada code editor

```
c(1:5)
```

Klik tombol **Run** dan perhatikan output yang dihasilkan pada bagian **Console** yang ditampilkan sebagai berikut.

```
> c(1:5)
[1] 1 2 3 4 5
```

Terlihat bahwa perintah `c(1:5)` membuat vector dengan lima urutan angka yang dimulai dari 1 dan diakhiri nilai 5.

```
c(1:5)
## [1] 1 2 3 4 5
```

Menggunakan Fungsi Summary

Kekuatan bahasa pemrograman R terdapat pada fungsi-fungsi bawaan yang kaya untuk dipergunakan dalam analisis data. Salah satu fungsi ini adalah fungsi bernama **summary** yang bisa digunakan untuk menyimpulkan data yang lagi kita proses. Kesimpulan ini berupa nilai statistik deskriptif dari data yang dimiliki, diantaranya nilai minimum, nilai kuartil pertama, nilai tengah (median), nilai kuartil ketiga, dan nilai maksimum data.

Ketikkanlah perintah berikut di Code Editor agar kita dapat melihat karakteristik dari vector `c(1:5)`

```
summary(c(1:5))
```

Jalankan dan kamu akan mendapatkan hasil pada Console seperti berikut. Kamu dapat *men-scroll* ke bawah untuk mengetahui penjelasannya.

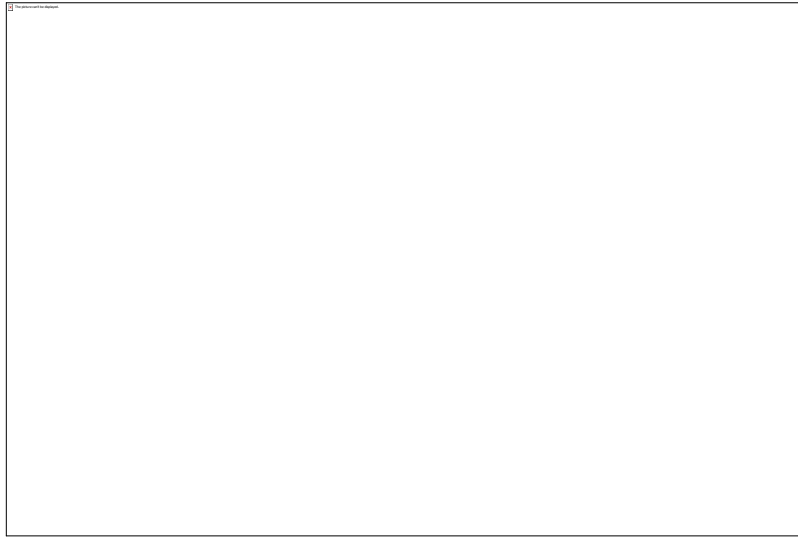
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1	2	3	3	4	5

Ini artinya dari vector tersebut terdapat angka paling kecil 1 (Min), angka paling besar 5 (Max), angka rata-rata 3 (Mean), dan angka tengah 3 (Median). Untuk 1st Qu dan 3rd Qu kita abaikan dulu.

Klik tombol **Submit** Code untuk melanjutkan ke bagian akhir pelajaran kita.

```
summary(c(1:5))
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##      1      2      3      3      4      5
```

Selamat untuk kamu!



Kamu baru saja mempelajari dasar awal Bahasa Pemrograman R sebagai langkah awal menjadi seorang Data Scientist.

Masih banyak modul pelajaran yang akan kamu pelajari lebih dalam untuk mengaplikasikan teknik data science secara tepat dengan menggunakan beragam studi kasus Industri.