

Jawaban

1. Apa itu *Data Science* ?

Data Science merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data yang sangat besar dengan menggunakan metode-metode yang berisikan dari *Artificial Intelligence*, *Machine Learning*, Statistik, dan *Database System*.

2. Apakah dibutuhkan jurusan/bidang keilmuan tertentu untuk mendalami *Data Science* ?

Secara garis besar tentu dibutuhkan. Misalnya seperti ilmu statistika, ilmu komputer, dan ilmu bisnis. Ketiga bidang ilmu tersebut akan sangat membantu seseorang dalam mendalami *Data Science*. Namun untuk seseorang yang berangkat dari latar belakang pendidikan/jurusan lain tidak menutup kemungkinan untuk bisa mendalami *Data Science* dengan mempelajari *skill-skill* yang dibutuhkan.

3. *Skill* apa saja yang dibutuhkan untuk menjadi seorang *Data Scientist* ?

Skill yang dibutuhkan untuk menjadi *Data Scientist* dibagi menjadi dua, yaitu skill teknis dan non-teknis. Skill teknis meliputi *statistical analytic* dan keterampilan dalam menggunakan *cloud computing* seperti *data mining* dan *processing* yang mengekstrak informasi dari sekumpulan data yang terstruktur maupun tidak terstruktur.

Sementara skill non-teknis yaitu memiliki pengetahuan yang luas tentang bisnis dan memiliki keterampilan komunikasi yang kuat. Seorang *Data Scientist* harus memahami data lebih baik dari siapa pun. Namun, agar berhasil dalam perannya membantu organisasi bisnis, ia harus berhasil mengkomunikasikan pemahamannya dengan orang yang akan menggunakan data tersebut yang tentunya mereka adalah orang non-teknis yang tidak memiliki pengetahuan tentang data.

4. Apa itu *Data Cleaning* ?

Data Cleaning adalah salah satu tahapan dalam *Data Preparation*. Proses *Data Cleaning* merupakan proses yang penting dilakukan karena akan mempengaruhi hasil *modelling Machine Learning*. Hampir semua *dataset* yang akan digunakan untuk *modelling* tidak dalam kondisi siap pakai, yang artinya harus dibersihkan dan diperbaiki lebih dulu. Data yang dibersihkan tersebut adalah data yang salah, rusak, tidak akurat, tidak lengkap dan salah format.

5. Jelaskan Metode Statistika yang sangat bermanfaat dalam *Data Science* !

Metode yang sangat bermanfaat adalah regres. Regresi adalah salah satu metode pengolahan data statistik yang berguna untuk menentukan hubungan sebab-akibat antara variabel dengan variabel lainnya. Di dalam analisis regresi, hubungan antara variabel bersifat linier, dimana perubahan pada variabel X akan diikuti oleh perubahan pada variabel Y secara tetap. Sedangkan dalam hubungan yang non linier, perubahan X tidak diikuti oleh variabel Y secara proporsional.

6. Apa Itu *Scatter Plot* ? Untuk apa *Scatter Plot* digunakan ?

Scatter plot adalah sebuah grafik yang biasa digunakan untuk melihat suatu pola hubungan antara 2 variabel. *Scatter Plot* atau diagram pencar merupakan diagram yang digunakan untuk menampilkan sebaran data. Seringkali dipakai untuk melakukan analisa regresi dan korelasi, runtun waktu (*time series*), bahkan dapat digunakan untuk melakukan *statistical learning*, salah satu topik populer dalam *Machine Learning*.

7. Jelaskan perbedaan *Supervised* dan *Unsupervised Learning* !

Perbedaan yang mencolok dari keduanya adalah pada saat melatih model *Machine Learning* apakah menggunakan data label atau tidak. Pendekatan *Supervised Learning* mempunyai input dan output yang jelas sehingga mampu melakukan prediksi dan klasifikasi berdasarkan data yang telah ada sebelumnya. Sementara pendekatan *Unsupervised Learning* memodelkan sekumpulan input secara otomatis tanpa adanya panduan output yang diinginkan.

Teks: Trailer Spiderman No Way Home	>	Topik: Film
Teks: Coldplay X BTS – My Universe	>	Topik: Music
Teks: Squid Game di Roblox!	>	Topik: Game

Sebagai contoh, jika data pelatihan yang dimiliki adalah pasangan-pasangan teks dan topiknya seperti di atas, dimana teks adalah input dan topik adalah outputnya. Maka algoritma pelatihan yang memanfaatkan label topik di atas disebut sebagai *Supervised Learning*. Masih data pelatihan yang sama, jika algoritma pelatihan tidak memanfaatkan label topik maka pembelajaran tersebut disebut *Unsupervised Learning*.

Studi Kasus

Prepare > Python > Introduction > Say "Hello, World!" With Python

Say "Hello, World!" With Python ★

30 more points to get your first star!

Rank: 1262925 | Points: 5/35



Problem Submissions Leaderboard Discussions Editorial Tutorial

You made this submission 2 minutes ago.

Score: 5.00 Status: **Accepted**

People who solved Say "Hello, World!" With Python attempted this next:

Python If-Else

Practice using if-else conditional statements

Solve Challenge

Submitted Code

Language: Python 3

Open in editor

```
1 print("Hello, World!")
2
```

NEED HELP?

- View tutorial
- View discussions
- View editorial
- View top submissions

You made this submission a few seconds ago.

Score: 10.00 Status: **Accepted**

People who solved Python If-Else attempted this next:

Arithmetic Operators

Addition, subtraction and multiplication.

Solve Challenge

Submitted Code

Language: Python 3

Open in editor

```
1 #!/bin/python3
2
3 import math
4 import os
5 import random
6 import re
7 import sys
8
9
10
11 if __name__ == '__main__':
12     n = int(input().strip())
13     if n%2 or 6 <= n <= 20:
14         print('Weird')
15     else:
16         print('Not Weird')
```

[Problem](#)[Submissions](#)[Leaderboard](#)[Discussions](#)[Editorial](#)[Tutorial](#)

You made this submission a few seconds ago.

Score: 10.00 Status: **Accepted**

People who solved **Arithmetic Operators** attempted this next:

Python: Division

Division using `__future__` module.

[Solve Challenge](#)

Submitted Code

Language: Python 3

[Open in editor](#)

```
1 if __name__ == '__main__':
2     a = int(input())
3     b = int(input())
4     print(a+b)
5     print(a-b)
6     print(a*b)
7
```

[Problem](#)[Submissions](#)[Leaderboard](#)[Discussions](#)[Editorial](#)[Tutorial](#)

You made this submission a few seconds ago.

Score: 10.00 Status: **Accepted**

People who solved **Python: Division** attempted this next:

Loops

Practice using "for" and "while" loops in Python.

[Solve Challenge](#)

Submitted Code

Language: Python 3

[Open in editor](#)

```
1 if __name__ == '__main__':
2     a = int(input())
3     b = int(input())
4     print(a//b)
5     print(a/b)
6
```

You made this submission a few seconds ago.

Score: 10.00 Status: **Accepted**

People who solved **Lists** attempted this next:

Tuples

Learn about tuples and compute hash(T).

[Solve Challenge](#)

Submitted Code

Language: Python 3

[Open in editor](#)

```
1 list1 = []
2 n = int(input())
3 for i in range(n):
4     s, *d = input().split()
5     d = list(map(int, d))
6     if s == "print":
7         print(list1)
8     else:
9         getattr(list1, s)(*d)
10
```

Problem

Submissions

Leaderboard

Discussions

Editorial

Tutorial

You made this submission a few seconds ago.

Score: 20.00 Status: **Accepted**

People who solved **Print Function** attempted this next:

Exceptions

Handle errors detected during execution.

[Solve Challenge](#)

Submitted Code

Language: Python 3

[Open in editor](#)

```
1 n = int(input())
2 print(''.join(map(str, range(1, n+1))))
3
```

You made this submission a few seconds ago.

Score: 20.00 Status: **Accepted**

People who solved **Map and Lambda Function** attempted this next:

Validating Email Addresses With a Filter

This question covers the concept of filters.

[Solve Challenge](#)

Submitted Code

Language: Python 3

[Open in editor](#)

```
1 cube = lambda x: x**3 # complete the lambda function
2
3 def fibonacci(n):
4     initiallist = []
5     for i in range(n):
6         if i < 2:
7             initiallist += [i]
8         else:
9             initiallist += [initiallist[-1] + initiallist[-2]]
10    return initiallist
11
12
```