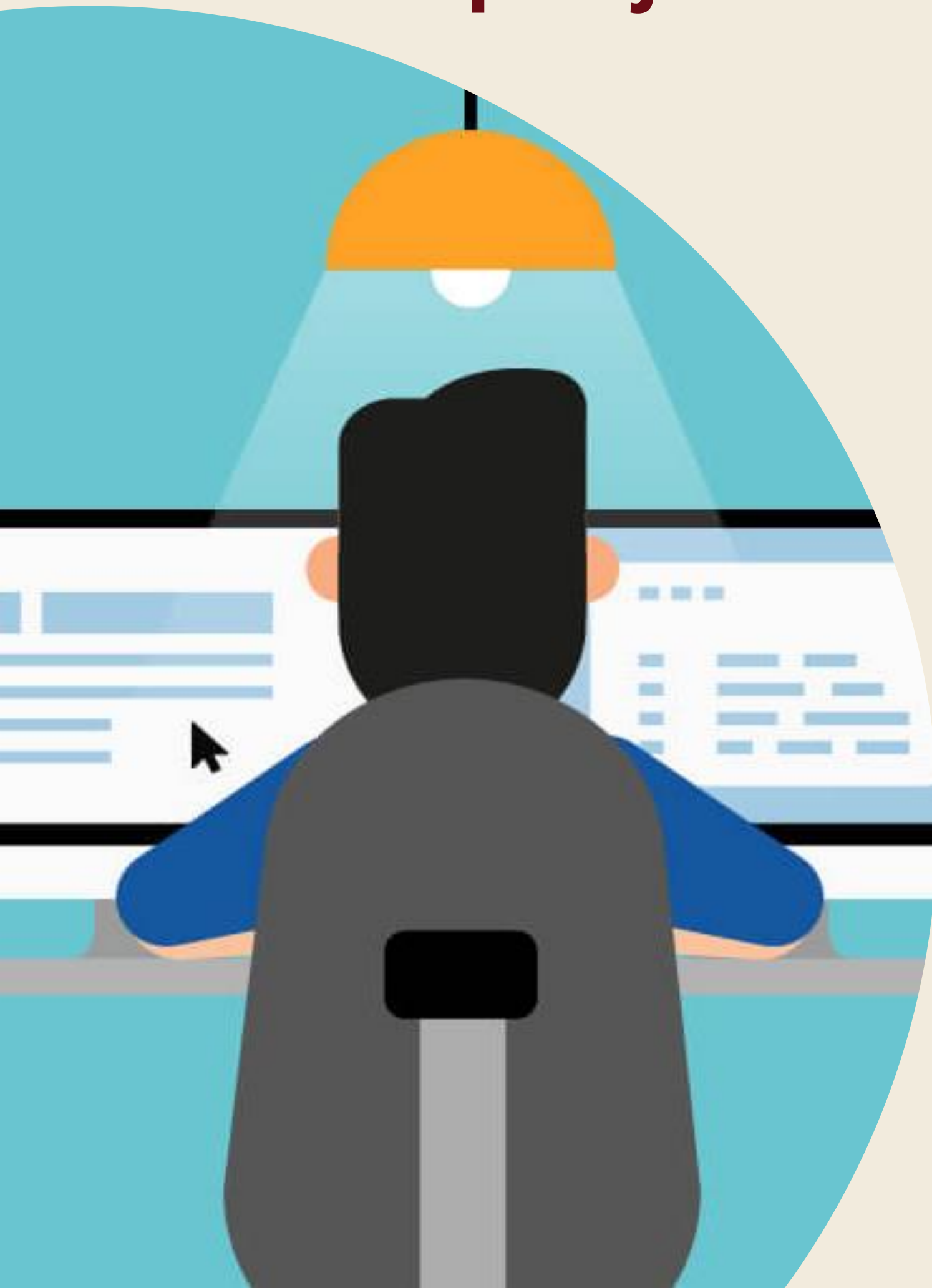DQLab

X

CTCORP

PROJECT DATA SCIENCE CHALLENGE WITH PYTHON

By Muhammad Rizky Halawi

ABOUT THIS PROJECT

DQLab

This a collaboration project with DQLab and CT-Corp. in this project I will do some skill test. That is:

1 **Probability Test**

2 **Statistics Test**

3 **Interpretation Models**

4 **Programming test**

Python

Python is programming language for **general purpose programming** and in **high- level programming language**.

As a general-purporse programming language, pythom can be use for some problem like web development, application mobile, **data science**, and etc.

Python get into high-level programming language because python's language is easy to read and write for human.

Programming python language found by **Guido Van Rossum** and introduced at 1991 as open-source project. License from python is an open source from python. that's mean everyone can develop computer program with python for the commercial purpose or not.

Question 1: Rugby Competition

In a competition have 23 teams rugby .
How much match must be played until we know the who's the Champion?

```
In [35]: tim_rugby = 23
         pertandingan = x
         x= tim_rugby - 1
         print(x)

         22
```

Note by my self:
If the totals team is 23 maybe 22 match will be played. Because 1 match can involve by 2 teams.

Question :Descriptive Statistic

From that question for all answer is true because mean is the central from data and connected by random variable and standart deviation to measure the amount of data distribution

Statistik Deskriptif

Manakah pernyataan yang benar dari pilihan di bawah ini

- ☐ Rata-rata adalah ukuran tendensi pusat suatu data
- ☐ Rata-rata empiris berhubungan dengan pemusatan variabel acak
- ☐ Standar deviasi empiris mengukur besarnya sebaran data
- ☑ Semua benar

Submit Answer

Question :Descriptive Statistic

The answers this problem is "K Means Clustering" is one of the "unsupervised machine learning algorithms" which is to find group in data, by number of group are represented by the totals clusters desired

Algoritma mana ya?

Setelah mengkaji perilaku sebuah populasi, Anda telah mengidentifikasi empat jenis individu spesifik yang bernilai pada kajian Anda. Anda akan menentukan semua pengguna yang paling serupa untuk setiap jenis individu itu. Algoritma manakah yang paling cocok untuk kajian ini?

Pilihlah satu jawaban yang benar.

- ☑ K-means clustering
- ☐ Linear regression
- ☐ Association rules
- ☐ Decision trees

Submit Answer

D Φ Lab

QUESTION: Binary Classification

The answers 1 and 4 it's correct because they have high accuracy and high positive rating true, which is 0.91 and 0.95. accuracy means true prediction ratio (positive or negative) with the whole data. the true positive rating is the rate at which an event actually occurs.

Binary Classification

Jika Anda bekerja pada suatu proyek yang menggunakan permasalahan *binary classification*. Kemudian, Anda mentraining sebuah model pada *training dataset* dan kemudian memperoleh *confusion matrix* setelah menerapkannya pada *validation dataset*, berikut ini

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Berdasarkan pada confusion matrix tersebut, pilihlah jawaban berikut ini mana yang memberikan prediksi yang benar?

1. *Accuracy* adalah ~0.91
2. *Misclassification rate* adalah ~ 0.91
3. *False-positive rate* adalah ~0.95
4. *True positive rate* adalah ~0.95

☐ 1 dan 3
☐ 2 dan 4
☑ 1 dan 4
☐ 2 dan 3

Submit Answer

7

Dict Data Type

Dict data type!

Tuliskanlah kode python di code editor yang akan menyatukan tiga dictionary berikut

```
dic1 = {1:10, 2:20}
dic2 = {3:30, 4:40}
dic3 = {5:50, 6:60}
```

menjadi satu dictionary dalam *dic4*

Gantilah _ _ _ dengan fungsi/literal/variabel yang sesuai

Result

```
In [1]: #DICT DATA TYPE
        dic1 = {1:10, 2:20}
        dic2 = {3:30, 4:40}
        dic3 = {5:50, 6:60}
        dic4 = {}
        for x in (dic1, dic2, dic3):
            dic4.update(x)
        print(dic4)

        {1: 10, 2: 20, 3: 30, 4: 40, 5: 50, 6: 60}
```

for combine 2 dictionaries or more can use "for" function. like above and use .update () for update the value in dictionary type

Soal: Fibonacci Sequence

Jumlah 7 Deret Pertama Fibonacci

Tulislah sebuah fungsi Python yang digunakan untuk menghitung n deret pertama Fibonacci yang dimulai dari 0.

```
Input  : n = 7
Output :  33
```

Gantilah _ _ _ dengan fungsi/literal/variabel yang sesuai

Result

```python
In [18]:  # Buat fungsi penjumlahan deret Fibonacci
          def calculateSum(n):
              if n<=0:
                  return 0
              fibo = [0] * (n+1)
              fibo[1] = 1
          # Initialisasi hasil ke dalam variabel sm
              sm = fibo[0] + fibo[1]
          # Tambahkan suku-suku berikutnya
              for i in range(2,n+1):
                  fibo[i] = fibo[i-1] + fibo[i-2]
                  sm += fibo[i]
              return sm
          # Evaluasi hasil deret untuk n = 7
          print(calculateSum(7))

          33
```

Notes:
Fibonacci formula = [0]* (n+1)

CONCLUSION

by completing this project, based on an assessment from DQLabs I've shown that I have:
- Ability to conceptualize knowledge in the probability field
- ability to understand the basics of statistics
- The ability to provide insight / decision making and support models built through the application of data science
-applying logical thinking through the code program that is asked using python

THANKYOU

SEE YOU ON NEXT PROJECT

CONTACT ME

YOUR SUBTITLE HERE

DRAMAGA, BOGOR, INDONESIA

https://www.kaggle.com/rizkyhalawi
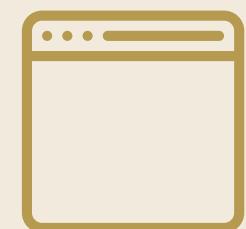
+62812131463091

rizkyhalawi@gmail.com

Rizky Halawi

Data Mining and Analyst at IPB University
East Jakarta, Jakarta, Indonesia
Joined 4 days ago · last seen in the past day

Competitions   Datasets   Notebooks (1)   Discussion   •••   Edit

titions   Datasets   Notebooks   Discussion
                       Novice   Novice   Novice

Unranked   Unranked   Unranked   Unranked