Nama : Yoel Oscar Werinussa
Nrp : 1672063
UTS : Pencarian Informasi Media Online

Jawaban.

1.

A. Distribusi Pareto adalah merupakan sebuah istilah yang diambil dari seorang teknik sipil, ahli ekonomi dan sosiologis asal italia yaitum *Vilfredo Pareto* , dan dapat disebut sebagai power law probability distribution dan telah di gunakan sebagai deskripsi dari social, kontrol kualitas, saintis, dan geofisikal. Setelah itu, *Vilfredo Pareto* melakukan observasi terhadap populasi pada 80% tanah di italia, yang mana hanya berpopulasi sekitar 20% saja, yang mana dilihat sebagai jumlah yang sangat sedikit dari 80% tersebut. Dan observasi itu dinamakan *Pareto Principle*

B. Distribusi Pareto berpotensi terjadi di dalam sosial media dikarenakan
Distribusi aktivitas pengguna tersebut tidak berubah dari waktu ke waktu, Dan setiap user selalu melakukan pengeditan yang lebih proporsional dengan suatu hal yang mereka angkat, dan itu semua terlepas dari seberapa aktif mereka.

Contoh yang dapat diambil dari distribusi pareto ialah,

Katakanlah, kita diperhadapkan dengan suatu platform tanya jawab atau biasa di sebut forum diskusi, jika kita melihat dari prinsip pareto, dan kita memakai teknik 80/20, maka bilamana suatu user memberikan suatu jawaban atau respon terhadap suatu pertanyaan yang di ajukan oleh user lainnya, maka jikalau petanyaan tersebut memiliki bobot atau kualitas yang baik, mungkin bisa saja jawaban yang diberikan tersebut menjadi jawaban bagi pertanyaan user lainnya, atau jika kita memiliki kasus lain, seperti platform yang menampung user nya yang dapat memberikan informasi yang mungkin sedang ramai-ramainya dibicarakan, seperti pada platform *twitter*. Pada platform tersebut, user dapat melihat 20% informasi yang mana dapat memenuhi 80% keingin tahuannya akan informasi yang sedang trending tersebut.

2.

A. Polya Urn atau biasa disebut Kendi Polya adalah merupakan tipe dari statistika yang dinamai dari *George Polya* , didalam kendi polya, kita dapat melihat beberapa objek yang di presentasikan sebagai suatu bola berwarna didalam sebuah guci atau kendi, dan pada kendi polya secara dasar, hanya di dapati dua objek saja yaitu bola x putih dan bola y hitam. Dan memiliki sifat ketika sebuah bola hitam dilempar keluar dari sebuah kendi dan dig anti, maka beberapa bola dengan warna yang sama di tambahkan.

B.

Contoh yang dapat diambil dari Kendi Polya ialah,

Katakanlah kita sedang dalam platform dimana user dapat berinteraksi atau mengajak user lainnya, seperti twitter, katakanlah bola 1 sebagai bola hitam, yang mana bola ini ialah user dan bola 2 sebagai bola putih yang mana bola ini ialah konten, pada sewaktu waktu, bola hitam masuk kedalam kendi dimana berisi 1 bola putih, dan ketika bola itu dikeluarkan, maka ia akan membawa bola hitam lainnya, yang mana kita mempresentasikannya sebagai suatu konten, yang di akses oleh beberapa user didalam twitter, dan seketika, user memiliki hubungan satu dengan yang lainnya, seperti melakukan saling tweet ulang atau bahkan melihat isi detail dari profil user lainnya didalam satu kendi.

3. A.

Term(t)	D1	D2	D3	D4	D5
auto	1	0	2	0	0
car	0	1	1	0	0
wash	0	1	1	0	1
machine	0	0	0	1	1

IDF Jika rumusnya = idf = 1/df :

Term(t)	Df	Idf
auto	3	1 / 3 = 0.3
car	2	1 / 2 = 0.5
wash	3	1 / 3 = 0.3
machine	2	1 / 2 = 0.5

Jika rumusnya = idf = N/ df :

							TF.IDF				
Term(t)	D1	D2	D3	D4	D5	IDF	D1	D2	D3	D4	D5
							0.2218	0	0.4436	0	0
auto	1	0	2	0	0	Log(5/3) = 0.2218					
							0	0.3979	0.3979	0	0
car	0	1	1	0	0	Log(5/2) = 0.3979					
							0	0.2218	0.2218	0	0.2218
wash	0	1	1	0	1	Log(5/3) = 0.2218					
							0	0	0	0.3979	0.3979
machine	0	0	0	1	1	Log(5/2) = 0.3979					

B. Jika ada kueri ‘car wash’ berikan urutan skor dokumen yang paling relevan dengan cosine similarity

Length 1 = sqrt(0.2218^2 + 0.4436^2) = 0.4959
Length 2 = sqrt(0.3979^2 + 0.3979^2) = 0.5627
Length 3 = sqrt(0.2218^2 + 0.2218^2 + 02218^2) = 2218
Length 4 = sqrt(0.3979^2 + 0.3979^2) = 0.5627
Length Q = sqrt(0.1989^2 + 0.1109^2) = 0.2277

CosSim(1,Q) = (0.1989 * 0 + 0 * 0 + 0.1989 * 0 + 0 * 0 + 0 * 0) / (0.4959 * 0.2277) = 0
CosSim(2,Q) = (0 * 0 + 0.1989 * 0.1989 + 0.1989 * 0.1989 + 0 * 0 + 0 * 0) / (0.5627 * 0.2277) = 0.6175322116475
CosSim(3,Q) = (0 * 0 + 0.1989 * 0.1109 + 0.1989 * 0.1109 + 0 * 0 + 0.1989 * 0.1109) / (2218 * 0.2277) = 0.0001314317181
CosSim(4,Q) = (0 * 0 + 0 * 0 + 0 * 0 + 0.1989 * 0 + 0.1989 * 0) / (0.5627 * 0.2277) = 0

Jadi Skor yang paling relevan yaitu ada di CosSim(2,Q) ,dan CosSim(3,Q) dengan nilai :
CosSim(2,Q) = 0.6175322116475
CosSim(3,Q) = 0.0001314317181

4. Menerapkan ARIMA pada data set ‘question_philoit.csv’

Pertama, saya melakukan import terlebih dahulu pandas, yang merupakan package dari python,

```
import pandas as pd

df = pd.read_csv('question_philoit.csv')

print('Shape of data',df.shape)

df.head()

df
```

df.shape, berfungsi menampilkan bentuk dari data yaitu seperti dibawah berikut
Shape of data (91163, 15)
Lalu saya mencoba melakukan sorting beberapa kolom dan baris untuk menampilkan data dari bulan Februari hingga akhir april dengan cara memilah baris sesuai dengan index dimana terdapat bulan Februari dan april 2020

```
row_start = df.index[1704]
row_end = df.index[4927]

df.loc[row_start:row_end, 'created_at'].to_csv(r'E:\KULIAH\Materi Kuliah\Pencarian Informasi media O
nline\UTS\time.csv')
```

disini saya melakukan dua kali export, yang mana nantinya saya menggabungkan kedua file tersebut