

UTS

Data Science

Berita bohong atau hoaks adalah informasi yang sesungguhnya tidak benar, tetapi dibuat seolah-olah benar adanya. Hal ini tidak sama dengan rumor, ilmu semu, atau berita palsu, maupun April Mop. Tujuan dari berita bohong adalah membuat masyarakat merasa tidak aman, tidak nyaman, dan kebingungan. Dataset <https://drive.google.com/file/d/1zAbuH7nS57RssHLEd-sZoOm6R2wpvNQw/view?usp=sharing> merupakan salah satu dataset berita hoaks yang digunakan untuk klasifikasi dalam bahasa Indonesia.

In [2]:

pd.read_csv('Hoax_training.csv')

Out[2]:

	ID	label	tanggal	judul	narasi	nama file gambar
0	71	1	17-Aug-20	Pemakaian Masker Menyebabkan Penyakit Legionna...	A caller to a radio talk show recently shared ...	71.jpg
1	461	1	17-Jul-20	Instruksi Gubernur Jateng tentang penilangan ...	Yth.Seluruh Anggota Grup Sesuai Instruksi Gube...	461.png
2	495	1	13-Jul-20	Foto Jim Rohn: Jokowi adalah presiden terbaik ...	Jokowi adalah presiden terbaik dlm sejarah ban...	495.png
3	550	1	8-Jul-20	ini bukan politik, tapi kenyataan Pak Jokowi b...	Maaf Mas2 dan Mbak2, ini bukan politik, tapi k...	550.png
4	681	1	24-Jun-20	Foto Kadrun kalo lihat foto ini panas dingin	Kadrun kalo lihat foto ini panas dingin . .	681.jpg
...
4226	999287	1	17-Jan-16	: KPK Dilarang Membawa Brimob Bersenjata Masuk...	KPK Dilarang Membawa Brimob Bersenjata Masuk G...	999287.jpg
4227	999472	1	20-Aug-20	Foto pejabat keuangan dibawah palu arit	Jangan mau dialihkan kepada pakaian adat... FOKU...	999472.png
4228	999501	1	17-Aug-20	Gambar Denny Siregar Musuh Warga Tasikmalaya d...	Akhir nya bisa terkenal bang denny	999501.jpg
4229	999848	1	21-Jul-20	Kaesang: Bapak Saya dengan Kesederhaan Bisa Ni...	Bapak Saya dengan Kesederhaan Bisa Nipu Rakyat...	999848.jpg
4230	999860	1	20-Jul-20	Laser dari termometer gun akan merusak struktu...	IN: Saya nolak. Kalau Anda mau periksa, bukan ...	999860.png

4231 rows x 6 columns

Pertanyaan :

1. Deskripsikan dataset di atas dan analisis 6 atribut yang terdapat pada dataset tersebut! (10%)
2. Lakukan data preprocessing pada dataset di atas! (15%)
3. Kelompokkan jumlah kasus hoaks per bulan dan visualisasikan hasilnya! (10%)
4. Kelompokkan jumlah kasus hoaks per tahun dan visualisasikan hasilnya! (10%)
5. Hitung rata-rata, modus dan median dari data kelompok jumlah kasus hoaks per bulan! (20%)
6. Hitung varian, standar deviasi dari kelompok jumlah kasus hoaks per bulan! (15%)
7. Hitung nilai skewness dari dataset tersebut dan visualisasikan hasilnya! (10%)
8. Simpulkan wawasan apa yang bisa diperoleh dari dataset di atas! (10%)