

Menghitung Skewness dan Kurtosis dari Persebaran dan Distribusi Panjang Sepal Bunga Iris

Tessa Kania Sagala¹, Renisha Putri Giani², Rian Bintang Wijaya³, Gymnastiar Al Khoarizmy⁴,
Virdio Samuel Saragih⁵

Jurusan Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Lampung Selatan, Indonesia

Email: tessa.122450040@student.itera.ac.id, renisha.122450079@student.itera.ac.id,
rian.122450094@student.itera.ac.id, gymnastiar.122450096@student.itera.ac.id,
virdio.122450124@student.itera.ac.id

1. Pendahuluan

Distribusi data adalah sebuah fungsi yang menunjukkan semua nilai dari suatu data dan seberapa sering nilai tersebut terjadi (Nugroho et al., n.d., 2020). Statistika (Wicaksono et al., 2021) adalah cabang ilmu yang mempelajari tentang bagaimana mengumpulkan, menganalisis, dan menginterpretasikan data. Ukuran penyebaran data ada 2, yaitu kemiringan (skewness) dan keruncingan (kurtosis). Skewness menggambarkan asimetris distribusi data dan kurtosis mengukur tingkat keruncingan distribusi data. Dua ukuran ini merupakan sebuah bentuk dan sifat dari distribusi data yang tidak dapat diberikan oleh ukuran statistik deskriptif seperti mean atau median.

Kurtosis (Nugroho et al., 2020) adalah tingkat kepuncakan dari suatu distribusi data, biasanya diambil relatif terhadap distribusi normal. Derajat keruncingan distribusi data dapat dihitung berdasarkan rumus yang dispesifikasikan berdasarkan apakah data tersebut berkelompok atau tidak. Sifat sifat dapat ditentukan berdasarkan derajat keruncingannya, apabila derajat keruncingan memiliki nilai yang sama dengan 3 maka keruncingan data tersebut adalah mesokurtis, namun apabila lebih dari 3 maka keruncingan data tersebut adalah leptokurtis dan apabila keruncingan kurang dari 3 maka disebut platikurtis.

Skewness merupakan ukuran yang menggambarkan model distribusi yang memiliki derajat kemiringan tertentu. Nilai skewness dapat bersifat positif, negatif atau nol. Nilai skewness positif menunjukkan distribusi yang lebih condong ke kanan, sedangkan skewness negatif menunjukkan distribusi data yang lebih condong ke kiri dan nilai skewness nol menunjukkan bahwa distribusi data simetris.

Pemahaman mengenai skewness dan kurtosis memegang peran penting dalam berbagai bidang karena dengan memahami dua konsep ini, peneliti dan analis data dapat mengeksplor dan menggambarkan data dengan lebih baik. Data yang digunakan pada artikel kali ini adalah data bunga Iris. Data yang akan diteliti adalah data UCI machine learning yaitu data iris (Mutrofin et al., 2014, #). Data iris adalah data bunga iris yang terdiri dari 4 atribut, 3 kelas dan 150 data. Data tersebut berisi informasi panjang & lebar sepal, panjang & lebar petal dan class dari setiap 3 spesies bunga Iris (Iris setosa, Iris virginica dan Iris versicolor), tetapi khusus untuk percobaan skewness dan kurtosis kami mengambil data panjang sepal dalam satuan cm.

2. Metode

Dalam pengerjaan perhitungan skewness dan kurtosis kami menggunakan beberapa fungsi:

2.1. Fungsi mean()

Fungsi mean(data) digunakan untuk menghitung rata-rata dari data. Fungsi tersebut dilakukan dengan menjumlahkan semua elemen dan selanjutnya membagi dengan jumlah elemen dari data tersebut.

2.2. Fungsi variance()

Fungsi variance(data) digunakan untuk menghitung nilai varians. Varians merupakan ukuran dari seberapa jauh titik data tersebar dari nilai rata-rata. Untuk perhitungannya mengambil rata-rata kuadrat perbedaan antara setiap titik data dan rata-rata.

2.3. Fungsi standard_deviation()

Fungsi standard_deviation(data) menghitung simpangan baku dari data yang diberikan, yang dapat memberikan gambaran seberapa jauh titik data tersebar dari nilai rata-rata.

2.4. Fungsi skewness()

Fungsi skewness(data) digunakan untuk menghitung kemiringan data. Untuk nilai kemiringan positif menunjukkan distribusi yang lebih panjang ke sebelah kanan dari rata-rata, sedangkan kemiringan negatif menunjukkan distribusi yang lebih panjang ke sebelah kiri dari rata-rata.

2.5. Fungsi kurtosis()

Fungsi kurtosis(data) untuk menghitung kurtosis data. Kurtosis mengukur tingkat kecuraman puncak distribusi data. Nilai kurtosis yang tinggi memperlihatkan distribusi data yang memiliki bagian puncak yang tajam dan ekor yang tebal. Sedangkan untuk nilai kurtosis yang rendah dapat dilihat pada distribusi yang lebih datar jika dibandingkan dengan distribusi normal.

3. Pembahasan

Pada artikel ini kami menggunakan dataset berupa data sepal length, sepal width, petal length dan petal width dengan satuan cm. Dalam pembuatannya kami menggunakan beberapa fungsi yang dikombinasikan untuk mengetahui skewness dan kurtosis dari dataset tersebut, yaitu mendefinisikan fungsi mean, variance dan standard_deviation yang lebih lanjut digunakan pada perhitungan dalam fungsi skewness dan kurtosis, serta menggunakan dua library yaitu library Pandas dan library Sklearn. Library Pandas digunakan untuk membuat DataFrame kosong dan mengisi kolom DataFrame dengan nilai-nilai yang diimport dengan menggunakan library Sklearn.

3.1 Pemanggilan Data dengan Library Pandas dan Sklearn

```
from sklearn import datasets
#menggunakan datasets built-in sklearn

import pandas as pd
#mengimport library dataframe

datairis = pd.DataFrame(datasets.load_iris().data, columns=datasets.load_iris().feature_names)
#memuat dataset iris dari library sklearn
datairis
```

Pada perintah di atas, kode 'from sklearn import datasets' yang digunakan untuk mengimport data yang kita butuhkan dari library Sklearn yakni data *sepal length*, *sepal width*, *petal length* dan *petal width* dengan satuan cm.

Selain itu juga, kode 'import pandas as pd' kita gunakan untuk visualisasi data berupa dataframe yang disimpan dalam variabel data iris sehingga lebih mudah untuk pembacaan dari data yang kita import dari library sklearn tersebut, sehingga muncul tampilan sebagai berikut:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

3.2 Mendefinisikan berbagai fungsi yang digunakan

```
def mean(data):
    # Mencari rata-rata sebuah data dengan cara menjumlahkan data
    # menggunakan fungsi sum, dan dibagi dengan total data menggunakan fungsi len
    return sum(data) / len(data)

def variance(data):
    mu = mean(data)
    # Mencari rata-rata
    return sum((x - mu) ** 2 for x in data) / (len(data) - 1)

def standard_deviation(data):
    return variance(data) ** 0.5

def skewness(data):
    mu = mean(data)
    sigma = standard_deviation(data)

    #
    return sum((x - mu) ** 3 for x in data) / ((len(data) - 1) * sigma ** 3)

def kurtosis(data):
    mu = mean(data)
    sigma = standard_deviation(data)
    return sum((x - mu) ** 4 for x in data) / ((len(data) - 1) * sigma ** 4) - 3
```

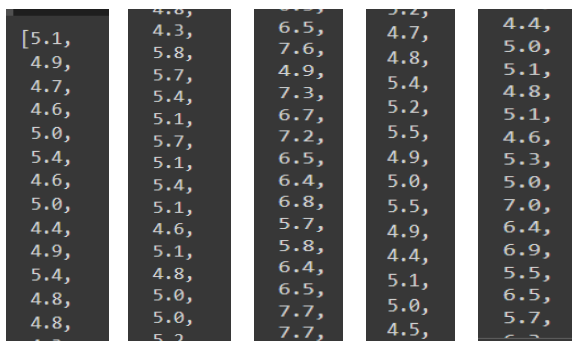
Pada tampilan baris kode di atas yakni adalah pendefinisian fungsi yakni fungsi mean untuk mencari rata-rata dari data, fungsi variance untuk mencari variansi data, fungsi standard_deviation untuk mencari standar deviasi, fungsi skewness untuk mencari nilai kemiringan dan juga fungsi kurtosis untuk mencari nilai keruncingan data. Fungsi yang utama kita gunakan adalah fungsi

skewness dan juga fungsi kurtosis yang akan mencari masing-masing nilai kemiringan dan keruncingan data, namun dalam fungsi skewness dan kurtosis membutuhkan perhitungan dari ketiga fungsi lainnya.

```
datalength = datairis['sepal length (cm)'].to_list()
# mengubah kolom panjang sepal dari dataframe menjadi dalam bentuk list
datalength
```

3.3 Mengubah data berupa DataFrame menjadi list

Baris kode di atas digunakan untuk mengubah kolom data datalength yang awalnya dari dataframe menjadi list yang akan digunakan secara lanjut dalam perhitungan skewness dan juga kurtosis pada baris kode selanjutnya, berikut adalah tampilan output dari baris kode di atas.



5.1,	4.0,	4.4,	5.2,	4.4,
4.9,	4.3,	6.5,	4.7,	5.0,
4.7,	5.8,	7.6,	4.8,	5.1,
4.6,	5.7,	4.9,	5.4,	4.8,
5.0,	5.4,	7.3,	5.2,	5.1,
5.4,	5.1,	6.7,	5.5,	4.6,
4.6,	5.7,	7.2,	4.9,	5.3,
5.0,	5.1,	6.5,	5.0,	5.0,
4.4,	5.4,	6.4,	5.5,	7.0,
4.9,	5.1,	6.8,	4.9,	6.4,
5.4,	4.6,	5.7,	4.4,	6.9,
4.8,	5.1,	5.8,	5.1,	5.5,
4.8,	4.8,	6.4,	5.0,	6.5,
4.3,	5.0,	6.5,	4.5,	5.7,
4.7,	5.2,	7.7,	4.5,	6.3,

3.4 Mencari nilai skewness dan kurtosis

```
print(skewness(datalength))
print(kurtosis(datalength))
```

```
0.31071214388181073
-0.5897441625988109
```

Baris kode tersebut untuk mencari nilai kemiringan dan keruncingan agar dapat membandingkan penyebaran data dengan distribusi normal dengan memanggil fungsi skewness dan fungsi kurtosis yang telah didefinisikan sebelumnya dan variabel datalength sebagai datanya.

4. Kesimpulan

Berdasarkan pembahasan yang telah dijelaskan, konsep fungsi sebagai objek dalam bahasa pemrograman Python telah diimplementasikan untuk melakukan analisis statistik pada dataset dengan

menggunakan pendekatan pemrograman berbasis fungsi. Fungsi-fungsi dikembangkan untuk menghitung parameter-parameter penting seperti mean, varians, standar deviasi, skewness, dan kurtosis dengan memanfaatkan konsep fungsi sebagai objek yang memungkinkan fungsi diperlakukan sebagai nilai yang dapat diteruskan sebagai argumen ke fungsi lain atau dioperasikan dengan operasi aritmatika dan logika. Pendekatan ini memfasilitasi pembuatan kode yang modular, efisien, dan mudah dipertahankan. Sebagai studi kasus, fungsi-fungsi tersebut diaplikasikan pada dataset iris untuk menganalisis distribusi fitur 'sepal length (cm)', dengan hasil menunjukkan data panjang sepal memiliki sedikit kemiringan ke kanan (skewness positif 0,31) dan distribusi sedikit lebih datar (kurtosis negatif -0,59) dibandingkan distribusi normal, mengilustrasikan bagaimana fungsi sebagai objek dapat dimanfaatkan untuk mengekstrak wawasan berharga dari data melalui pemrograman berbasis fungsi.

5. Referensi

- Mauldina, A. (2021). Menentukan Koefisien Kurtosis. *Volume 5 Nomor 3*, 114576-114579.
- Mutrofin, S., Izzah, A., Kurniawardhani, A., & Asrur, M. (2014, September). Optimasi Teknik Klasifikasi Modified K Nearest Neighbor Menggunakan Algoritma Genetika. *Jurnal Gamma, Issn 0216-9037, Volume 10, Nomor 1*, 130-134.
- Nugroho, D. A., Firdaus, M., & Ariyanto, H. D. (n.d.). Analisis Ukuran Penyebaran Data (Kemiringan dan Keruncingan)(Studi Kasus:Riwayat Penjualan Usaha Makanan Ibu Apri).
- Setiawan, A. (n.d.). Penentuan Distribusi Skewness Dan Kurtosis Dengan Metode Resampling Berdasar Densitas Kernel (Studi Kasus Pada Analisis Inflasi Bulanan Komoditas Bawang Merah, Daging Ayam Ras Dan Minyak Goreng Di Kota Semarang). *Prosiding Seminar Nasional Sains Dan Pendidikan Sains Vii Uksw*, 240-247.