

Standar Deviasi dan Variabilitas Hujan Antar Waktu

Dalam analisis hujan, standar deviasi dan variabilitas antar waktu digunakan untuk memahami sejauh mana nilai curah hujan berubah dari rata-rata dalam periode tertentu. Standar Deviasi dan Variabilitas cocok digunakan jika:

- Ingin melihat fluktuasi hujan dalam window kecil (60 menit).
- Tujuan analisis adalah memonitor perubahan curah hujan dalam jangka pendek.
- Ada kebutuhan untuk mendeteksi pola stabil atau ekstrem dalam curah hujan per window.

Standar Deviasi dan Variabilitas kurang cocok digunakan jika:

- Perubahan curah hujan sangat kecil atau mendekati nol dalam window 60 menit.
- Data sangat bervariasi dan terlalu pendek untuk mendapatkan tren yang berarti.
- Ingin mendeteksi pola jangka panjang (standar deviasi lebih stabil pada data yang lebih panjang).

1. Standar Deviasi (SD)

- Standar Deviasi mengukur seberapa jauh nilai individu dalam dataset tersebar dari rata-rata (mean).
- Jika SD kecil, data cenderung lebih terkonsentrasi di sekitar mean. Jika SD besar, data lebih tersebar.
- Semakin besar standar deviasi, semakin tinggi fluktuasi hujan antar periode.
- Formula standar deviasi untuk dataset curah hujan:

$$SD = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$$

di mana:

X_i adalah nilai curah hujan pada waktu ke- i ,

\bar{X} adalah rata-rata curah hujan,

N adalah jumlah data.

2. Variabilitas Hujan Antar Waktu

- Variabilitas hujan menggambarkan **sejauh mana curah hujan berubah** dari satu periode ke periode lain.
- Bisa dihitung dengan koefisien variasi (CV), yang dinormalisasi terhadap rata-rata:

$$CV = \frac{SD}{\bar{X}} \times 100\%$$

- Jika **CV tinggi**, berarti hujan sangat tidak stabil (misal: terjadi hujan ekstrem secara acak).
- Jika **CV rendah**, berarti pola hujan lebih stabil dan konsisten.

Menentukan Batas Berdasarkan Distribusi Normal

Jika data curah hujan berdistribusi normal, maka sekitar:

- 68% dari data berada dalam 1 SD dari mean ($\bar{X} \pm 1SD$)

- 95% dari data berada dalam 2 SD dari mean ($\bar{X} \pm 2SD$)
- 99,7% dari data berada dalam 3 SD dari mean ($\bar{X} \pm 3SD$)

Keterangan:

- 1 SD dari mean: mencakup sebagian besar data yang normal.
- 2 SD dari mean: mencakup lebih banyak data, tetapi ada kemungkinan nilai ekstrem.
- 3 SD dari mean: mencakup hampir semua data, di luar ini adalah anomali/outlier.

Jika data mengikuti distribusi normal, maka ada aturan yang disebut Empirical Rule (Aturan 68-95-99.7). Maksudnya adalah:

- 1 SD: sebagian besar data (68%) berada dalam $\pm 1 SD$ dari mean sehingga data masih dianggap wajar.
- 2 SD: 95% dari data berada dalam $\pm 2 SD$ dari mean sehingga data mulai menunjukkan perbedaan, tetapi masih dalam batas normal.
- 3 SD: 99.7% dari data berada dalam $\pm 3 SD$ dari mean sehingga jika ada data di luar rentang ini, maka itu kemungkinan besar outlier (anomali).

Jika curah hujan berada di luar batas ini, maka dianggap anomali atau nilai ekstrem.

- Jika nilai curah hujan lebih rendah dari batas rendah, kemungkinan terjadi anomali rendah (misalnya, hujan sangat ringan atau tidak ada hujan).
- Jika nilai curah hujan lebih tinggi dari batas tinggi, berarti curah hujan ekstrem (misalnya, hujan lebat atau badai).

Contoh perhitungan dalam Data Hujan.

Misalkan kita memiliki data curah hujan selama 10 hari (dalam mm):

Data Hujan = [10,25,30,5,50,15,20,40,35,28]

1. Hitung Mean (\bar{X}):

$$\bar{X} = \frac{10 + 25 + 30 + 5 + 50 + 15 + 20 + 40 + 35 + 28}{10} = 25.8 \text{ mm}$$

2. Hitung Standar Deviasi (SD):

$$SD = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$$

Hasilnya adalah:

$$SD \approx 13.13 \text{ mm}$$

3. Menentukan Rentang SD:

Batas	Perhitungan	Hasil	Keterangan
1 SD (68% data)	$25.8 \pm 1(13.13)$	[12.69mm, 38.93mm]	Data dalam rentang tersebut dianggap normal
2 SD (95% data)	$25.8 \pm 2(13.13)$	[-0.46 mm, 51,26 mm]	Data dalam rentang tersebut masih wajar, tetapi mulai ada nilai ekstrem

Commented [ID1]: SD dari residu (decomposition)

3 SD (99.7% data)	$25.8 \pm 3(13.13)$	[-13.59mm, 64.39mm]	Data di luar rentang tersebut dianggap anomali atau kejadian sangat jarang terjadi.
-------------------	---------------------	---------------------	---

Dalam contoh ini:

- 50 mm adalah mendekati batas 2 SD sehingga 50 mm berarti hujan cukup ekstrem.
- Tidak ada nilai di luar 3 SD, jadi tidak ada anomali ekstrem.
- Jika ada data di luar 2 SD bisa diartikan bahwa ada kejadian yang jarang terjadi sehingga perlu perhatian.

Menentukan Batas Menggunakan Koefisien Variasi (CV)

Koefisien Variasi (CV) melihat kestabilan data atau mengukur seberapa besar fluktuasi relatif terhadap rata-rata dalam bentuk persentase:

$$CV = \frac{SD}{\bar{X}} \times 100\%$$

- Jika $CV < 20\%$ maka Variabilitas rendah (hujan cukup stabil).
- Jika CV antara $20\% - 50\%$ maka Variabilitas sedang (hujan agak fluktuatif).
- Jika $CV > 50\%$ maka Variabilitas tinggi (hujan sangat tidak stabil).

Menentukan Boxplot (IQR) untuk Menentukan Outlier

Boxplot menggunakan Interquartile Range (IQR) untuk mendeteksi nilai ekstrem:

- $Batas\ bawah = Q1 - 1.5 \times IQR$
- $Batas\ atas = Q3 + 1.5 \times IQR$

Selain itu, alternatif yang dapat digunakan untuk analisis streaming data real-time, yaitu:

1. Moving Standard Deviation

- Menghitung standar deviasi secara dinamis berdasarkan window data terakhir.
- Membantu melihat bagaimana fluktuasi curah hujan berubah dari waktu ke waktu

2. Z-Score (Deteksi Anomali)

- Rumus Z-Score:

$$Z = \frac{X - \bar{X}}{SD}$$

- Jika nilai $Z > 2$ atau $Z < -2$, maka curah hujan dalam window dianggap anomali.

3. Exponential Moving Average (EMA)

- Memberikan bobot lebih besar pada data terbaru dibandingkan data lama dalam window.
- Lebih sensitive terhadap perubahan mendadak dibandingkan rata-rata biasa.

4. Autocorrelation

- Mengecek apakah ada pola berulang dalam hujan berdasarkan window data.
- Berguna untuk melihat apakah curah hujan saat ini dipengaruhi oleh hujan sebelumnya.

Standar deviasi dan variabilitas masih bisa digunakan, tetapi perlu diperhatikan bahwa window dengan panjang 6 data cukup pendek. Gunakan metode tambahan seperti moving standard deviation, Z-score, atau EMA untuk hasil lebih akurat. Jika data masuk secara live stream, bisa coba gunakan Pandas Rolling Window untuk analisis efisien.