

Pengaplikasian Outlier Detector

**Irhamna Mahdi¹, Salwa Farhanatussaidah², Ganiya Syazwa³, Reynaldi Rahmad⁴,
Syalaisha Andina Putriansyah⁵**

Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera

Email : irhamna.122450049@student.itera.ac.id¹, salwa.122450055@student.itera.ac.id²,
ganiya.122450073@student.itera.ac.id³, reynaldi.122450088@student.itera.ac.id⁴,
syalaisha.122450121@student.itera.ac.id⁵

1. Pendahuluan

Dalam kemajuan zaman, data menjadi salah satu aset penting bagi banyak orang baik organisasi dan individu. Akan tetapi, didalam Kumpulan data, sering terdapat titik yang tidak normal yang disebut “*outlier*”. Adanya *outlier* yaitu untuk menjadi sinyal penting terhadap kejadian yang tidak biasa dalam data, yang dapat mendeteksi atau mengindikasi masalah pada saat pengumpulan atau pemrosesan suatu data.

Outlier detector memiliki tujuan yaitu untuk menemukan pola yang tidak sesuai dengan perilaku yang diinginkan di dalam data. *Outlier detector* menjadi topik penting dalam penelitian diberbagai bidang ilmu, diantaranya ilmu statistik, data, dan mesin. Dalam ilmu statistik pada abad 19 setidaknya terdapat 4444 penelitian tentang *Outlier*. Sejak saat itu peneliti dan komunitas riset yang mengembangkan banyak metode tentang *Outlier Detector*.

2. Metode

2.1 Metode Z-Score

Metode Z-Score merupakan salah satu teknik yang biasa digunakan untuk mendeteksi outlier dalam suatu data. Z-Score juga dapat mengukur seberapa jauh data menyimpang dari rata-rata dalam hal standar deviasi [Anissa, N. (2017)]. Ketika akan menggunakan Z-Score, pertama kita akan menghitung nilai rata-rata dan standar deviasi dari data, lalu Z-Score untuk setiap titik akan dihitung menggunakan persamaan berikut:

$$z = (x - \mu) / \sigma$$

dimana:

- X = nilai data

- μ = rata- rata data
- σ = standar deviasi data

2.2 Metode Modified Z-Score

Modified Z-Score merupakan metode Z-Score yang dimodifikasi. Modifikasi ini dirancang untuk mengatasi beberapa kelemahan yang ada pada metode Z-Score asli, salah satunya dalam menangani data yang memiliki outlier ekstrim atau suatu distribusi yang tidak normal. Dibanding menggunakan rata-rata standar deviasi, metode ini akan menggunakan dan Median Absolute Deviation (MAD). [Winarso, E., & Edison, T. A. (2019)] Untuk mencari modified Z-Score digunakan persamaan berikut:

$$MZ = 0.6745(x_i - \tilde{x}) / MAD$$

dimana:

- x_i = nilai data
- \tilde{x} = median dari data
- MAD = Median Absolute Deviation dari data.
- Faktor 0.6745 = digunakan untuk menyesuaikan MAD (agar setara dengan standar deviasi jika data mengikuti distribusi normal)

2.3 Metode Percentile (IQR)

Metode Percentile (IQR) merupakan metode Interquartile Range (IQR), teknik deteksi outlier yang berdasarkan pada persentil data. Metode Percentile ini menggunakan kuartil data untuk menentukan batas-batas yang membantu mengidentifikasi outlier. Untuk menacari Percentile digunakan persamaan sebagai berikut:

$$IQR = Q3 - Q1$$

dimana:

- $Q3$ (Kuartil 3) = nilai data di bawah 25% dari data.
- $Q1$ (kuartil 1) = nilai data di bawah 75% dari data.

3. Pembahasan

3.1 Kode Program

```
def outlier_detector(method):
```

Membuat sebuah fungsi luar bernama outlier_detector, dimana fungsi ini menerima metode deteksi outlier sebagai argumen dan mengembalikan fungsi dektektor outlier berdasarkan metode yang dipilih. Fungsi outlier_detector memiliki tiga fungsi internal yang masing-masing menggunakan metode deteksi outlier yang berbeda, yaitu Z-Score, modified Z-Score, dan percentile. Ketiga fungsi internal tersebut menggunakan library NumPy untuk melakukan operasi matematika.

```
def z_score_outlier(data):
    import numpy as np
    threshold = 3
    mean = np.mean(data)
    std_dev = np.std(data)
    z_scores = [(x - mean) / std_dev for x in data]
    outliers = [data[i] for i in range(len(data)) if abs(z_scores[i]) > threshold]
    return outliers
```

Fungsi z_score_outlier(data) mengimplementasikan metode deteksi outlier menggunakan Z-Score. Z-Score untuk setiap titik data dihitung dengan membagi selisih antara nilai data dan rata-rata dengan standar deviasi. Nilai absolut dari Z-Score yang melebihi threshold akan diidentifikasi sebagai outlier.

```
def modified_z_score_outlier(data):
    import numpy as np
    threshold = 3.5
    median = np.median(data)
    median_absolute_deviation = np.median([np.abs(x - median) for x in
data])
    modified_z_scores = [0.6745 * (x - median) / median_absolute_deviation
for x in data]
    outliers = [data[i] for i in range(len(data)) if abs(modified_z_scores[i]) >
threshold]
    return outliers
```

Fungsi `modified_z_score_outlier(data)` merupakan fungsi yang didefinisikan untuk melakukan deteksi outlier menggunakan metode modified Z-Score. Fungsi ini menggunakan nilai median absolute deviation dalam menghitung nilai modified Z-Score dari setiap titik data yang kemudian digunakan untuk mengidentifikasi outlier berdasarkan threshold yang ditetapkan.

```
def percentile_outlier(data):
    import numpy as np
    q25, q75 = np.percentile(data, [25, 75])
    iqr = q75 - q25
    lower_bound = q25 - (1.5 * iqr)
    upper_bound = q75 + (1.5 * iqr)
    outliers = [x for x in data if x < lower_bound or x > upper_bound]
    return outliers
```

Fungsi `percentile_outlier(data)` merupakan fungsi yang mengimplementasikan metode deteksi outlier menggunakan Interquartile Range (IQR) atau percentile, dimana nilai data yang berada di luar rentang batas bawah ($Q1 - 1.5 * IQR$) dan batas atas ($Q3 + 1.5 * IQR$) akan diidentifikasi sebagai outlier.

```
def detector(data):
    if method == 'z_score':
        return z_score_outlier(data)
    elif method == 'modified_z_score':
        return modified_z_score_outlier(data)
    elif method == 'percentile':
        return percentile_outlier(data)
    else:
        raise ValueError("Invalid method. Available methods: 'z_score',
                          'modified_z_score', 'percentile'")

    return detector
```

Fungsi `detector(data)` merupakan fungsi closure yang memilih dan menjalankan fungsi deteksi outlier berdasarkan argumen 'method' yang diteruskan ke fungsi luar 'outlier_detector'. Jika metode yang diberikan tidak

valid, fungsi ini akan mengembalikan ValueError dengan pesan "Invalid method. Available methods: 'z_score', 'modified_z_score', 'percentile' ".

3.2 Hasil Program

kode di bawah merupakan contoh penggunaan dari fungsi outlier_detector yang telah dibuat, di mana data yang digunakan memiliki outlier jelas. Hasil deteksi outlier untuk masing-masing metode dicetak untuk ditampilkan kepada pengguna.

Input:

```
data = [1, 2, 3, 4, 5, 6, 7, 1000] # Data dengan outlier yang jelas
detect_outlier_z_score = outlier_detector('z_score')
detect_outlier_modified_z_score = outlier_detector('modified_z_score')
detect_outlier_percentile = outlier_detector('percentile')

print("Outlier detected using Z-Score method:",
detect_outlier_z_score(data))
print("Outlier detected using Modified Z-Score method:",
detect_outlier_modified_z_score(data))
print("Outlier detected using Percentile method:",
detect_outlier_percentile(data))
```

Output:

```
Outlier detected using Z-Score method: []
Outlier detected using Modified Z-Score method: [1000]
Outlier detected using Percentile method: [1000]
```

4. Kesimpulan

Outlier Detector adalah algoritma yang sering digunakan dalam bidang analisis data untuk mengidentifikasi banyaknya data secara teratur. *Outlier* bisa menjadi kesalahan yang terdapat pada pengukuran atau suatu kejadian langka yang tidak sewajarnya ada di dalam data. Dalam pengaplikasiannya digunakan Metode Z-Score untuk mengukur jarak data yang menyimpang dari rata rata dalam hal standar deviasi. Kemudian, digunakan juga Modifed Z-Score yang dimana Median Absolute Deviaton (MAD) digunakan untuk mengatasi data yang memiliki *outlier* ekstrim atau tidak normal.

Serta digunakan juga Metode Percentile dimana metode ini digunakan untuk menentukan batasan yang bisa membantu mengidentifikasi *Outlier*.

DAFTAR PUSTAKA

Anissa, N. (2017). Penggunaan Metode Z Score untuk Memprediksi Kemungkinan Kebangkrutan pada PT Mitra Adiperkasa Tbk. *Jurnal Ilmiah Ekonomi Bisnis*, 21(3).

Winarso, E., & Edison, T. A. (2019). Perbandingan Analisis Model Z'-Score Altman Modifikasi, Model X-Score Zmijewski, Model G-Score Grover, Dan Model S-Score Springate Untuk Menganalisis Ketepatan Prediksi Kebangkrutan:(Studi Pada Perusahaan Manufaktur Sektor Industri Pelengkap Otomotif Yang Terdaftar di BEI periode 2016-2017). *Journal of Accounting, Finance, Taxation, and Auditing (JAFTA)*, 1(2), 1-13.

Singh, K., & Upadhyaya, S. (2012). Outlier detection: applications and techniques. *International Journal of Computer Science Issues (IJCSI)*, 9(1), 307.