

NORMALISASI DATA

Febiya Jomy Pratiwi, Alvia Asrinda Br.Ginting
Farrel Julio Akbar, Akmal Faiz Abdillah
Daffa Ahmad Naufal

Program Studi Sains Data Institut Teknologi Sumatera
Jl. Terusan Ryacudu, Way Huwi, Kec. Jatiagung, Kabupaten Lampung Selatan
Lampung 35365

Email:

febiya.122450074@student.itera.ac.id, alvia.122450077@student.itera.ac.id
farrel.122450110@student.itera.ac.id, akmal.122450114@student.itera.ac.id
daffa.122450137@student.itera.ac.id

I. Pendahuluan

A. Latar Belakang

Di zaman digital seperti saat ini, pengolahan data sudah menjadi bagian yang tidak bisa dipisahkan dari kehidupan sehari-hari kita. Data yang dihasilkan dalam jumlah besar dari berbagai sumber, seperti aktivitas online, transaksi bisnis dan lain lain. Namun, data tersebut kerap tidak dapat dimanfaatkan secara efisien tanpa melalui tahapan normalisasi yang sesuai.

Proses normalisasi data adalah tahap yang sangat penting dalam mempersiapkan data sebelum menerapkan berbagai teknik analisis dan pembuatan model. Dalam kehidupan sehari-hari, ilustrasinya seperti ketika kita menyusun instrumen musik sebelum pertunjukan. Kita perlu memastikan bahwa setiap instrumen telah diperiksa, disetel dengan baik dan siap digunakan dengan maksimal.

Dengan cara yang sama, normalisasi data membantu mengubah data mentah menjadi format yang tepat dan siap digunakan untuk analisis lebih lanjut. Langkah ini melibatkan teknik-teknik standarisasi dan min-max scaling yang akan kita bahas lebih lanjut. Sebagai contoh sederhana, normalisasi data nilai siswa dapat membantu guru dalam membandingkan kinerja siswa dari berbagai ujian dengan skala yang seragam sehingga mempermudah dalam menilai kemajuan belajar siswa.

Kami akan membahas dua teknik normalisasi data, yaitu standarisasi dan min-max scaling, serta menerapkan keduanya dalam situasi data yang sebenarnya. Melalui pemahaman yang mendalam tentang tahapan normalisasi data, pembaca diharapkan dapat memperoleh wawasan lebih tentang pengolahan dan analisis data.

B. Tujuan

Adapun tujuan dari laporan ini adalah sebagai berikut, yaitu:

1. Memperkenalkan konsep normalisasi data
2. Mengaitkan konsep normalisasi data dengan kehidupan sehari-hari
3. Menguraikan tahapan normalisasi data

II. Metode

A. Normalisasi

Normalisasi merupakan teknik pra-pemrosesan yang berguna untuk mengubah data mentah menjadi data yang terstruktur dan meningkatkan efisiensi pembelajaran mesin. Ini berarti bahwa ketika data dikumpulkan dari berbagai sumber, data tersebut awalnya dalam bentuk mentah yang tidak dapat langsung digunakan untuk analisis atau pembelajaran mesin. [1]

B. Standarisasi Z-Score

Terdapat banyak metode data normalisasi, yang umum digunakan adalah metode Z-Score. Salah satu tujuannya adalah untuk menstandarisasi dataset. Nilai Z-Score merupakan ukuran penyimpangan data dari nilai rata-ratanya (μ) yang diukur dalam satuan standar deviasinya (σ). Z-Score ini juga disebut dengan Nilai Standar atau Nilai baku. [2]

C. Min-Max Normalization

Min-Max normalization adalah teknik normalisasi yang melibatkan transformasi linear terhadap data asli guna menciptakan keseimbangan nilai perbandingan antara data sebelum dan setelah proses. [3]

Dikarenakan metode normalisasi Min-Maks (0,1) mewakili semua fitur dalam rentang yang sama (0-1), metode yang diusulkan akan mengubah konsep ini dengan memberikan rentang yang berbeda untuk setiap fitur. Jika kita mengasumsikan f_1 sebagai fitur dengan korelasi 0, maka semua nilai akan memiliki nilai yang sama karena tidak ada korelasi dengan label kelas. Namun, fitur dengan korelasi 1 dengan label kelas akan mewakili nilai maksimum yang baru [4]

III. Pembahasan

A. Analisis Program

Untuk melakukan sebuah normalisasi data kita membutuhkan sebuah metode untuk menghasilkan hasil yang sesuai. Dalam mencari normalisasi data dibutuhkan metode, normalisasi, standarisasi Z-score, dan Min-Max Normalization. Teknik-teknik ini dapat digunakan untuk meningkatkan keterbacaan data dan menyederhanakan interpretasinya, menyelesaikan masalah terkait perbedaan skala antara fitur, dan memprosesnya menjadi model pembelajaran mesin menggunakan kedua pendekatan tersebut. Standardisasi dan penskalaan min-maks merupakan opsi yang harus dipertimbangkan bergantung pada persyaratan spesifik kumpulan data dan algoritme yang digunakan. Berikut code awal dalam melakukan normalisasi data.

```

1 import numpy as np
2
3 def standardize(X):
4     return (X - np.mean(X, axis=0)) / np.std(X, axis=0)
5 def min_max_scale(X):
6     return (X - np.min(X, axis=0)) / (np.max(X, axis=0) - np.min(X, axis=0))

```

Pengkodean ini merupakan implementasi dari dua fungsi untuk melakukan preprocess data pada suatu matriks menggunakan NumPy. Fungsi pertama, yaitu “standardize(X)”, fungsi ini bertujuan untuk menormalkan (normalisasi z-score) matriks masukan dengan nilai rata-rata setiap kolom pada kolom tersebut lalu bagi dengan simpangan baku kolom tersebut. Hal ini menghasilkan distribusi data dengan rata-rata 0 dan standar deviasi 1. Lalu selanjutnya fungsi “min_max_scale(X)”, proses standarisasi ini membantu algoritme pembelajaran mesin menyatu lebih cepat dan memperlakukan semua fitur pada skala yang sama. Fungsi ini dimaksudkan untuk melakukan penskalaan fitur (min-max scaling) pada matriks masukan X. Untuk menskalakannya, nilai minimum setiap kolom dibagi dengan selisih antara nilai maksimum dan terendahnya. Dengan menyesuaikan nilai antara 0 dan 1, rentang fitur akan diubah. Dengan memanfaatkan metode penskalaan ini, hubungan antar fitur dalam data dapat dipertahankan tanpa menimbulkan masalah dengan nilai outlier atau data yang berisi nilai berbeda. Selanjutnya code yang menampilkan dari normalisasi data berupa contoh.

```

1 # Buat dataset acak untuk keperluan contoh
2 np.random.seed(66)
3 dataset = np.random.rand(10, 3) # Contoh dataset 10 baris dan 3 kolom
4
5 print("Dataset Awal:")
6 print(dataset)
7
8 # Standardisasi dataset
9 standardized_data = standardize(dataset)
10 print("\nHasil Standardisasi:")
11 print(standardized_data)
12
13 # Min-Max Scaling dataset
14 scaled_data = min_max_scale(dataset)
15 print("\nHasil Min-Max Scaling:")
16 print(scaled_data)
17

```

Pertama, gunakan fungsi np.random.seed(66) untuk menentukan seed yang akan digunakan oleh generator nomor acak NumPy. Tujuannya adalah untuk menjamin bahwa data yang dihasilkan tetap tidak berubah selama setiap kali kode dijalankan. Hasilkan kumpulan data acak menggunakan np.random.rand(10, 3). Ini

akan membuat matriks 10x3 dengan nilai acak antara 0 dan 1. Cetak kumpulan data awal yang telah dibuat. Panggil fungsi `standardize()` untuk menormalkan kumpulan data yang dihasilkan. `standardize_data` adalah tempat hasilnya disimpan. Cetak hasil normalisasi kumpulan data. Untuk menskalakan kumpulan data ke ukuran maksimumnya, gunakan fungsi `min_max_scale()`. Variabel `scaled_data` berisi hasil. Cetak hasil penskalaan min-maks kumpulan data. Dalam contoh ini, kode menunjukkan cara menggunakan fungsi `Standardize()` dan `min_max_scale(boom)` untuk menghasilkan kumpulan data acak yang dinormalisasi dan kemudian diskalakan sesuai dengan prinsip min-max. Berikut output yang dihasilkan.

Dataset Awal:

```
[[0.15428758 0.13369956 0.36268547]
 [0.67910888 0.19445006 0.25121038]
 [0.75841639 0.55761859 0.51480292]
 [0.46779986 0.087176 0.82909544]
 [0.29864056 0.03134589 0.67800576]
 [0.90348903 0.51445118 0.53910547]
 [0.66432752 0.63405676 0.35341947]
 [0.02664251 0.16528984 0.87931899]
 [0.06782046 0.3690858 0.11550059]
 [0.09629414 0.0837697 0.08692703]]
```

Hasil Standardisasi:

```
[[ -0.8370447 -0.68002789 -0.37399998]
 [ 0.86966557 -0.39192791 -0.79803339]
 [ 1.12757232 1.33034353 0.2046305 ]
 [ 0.18249212 -0.90065877 1.40014905]
 [-0.36761117 -1.16542459 0.82542803]
 [ 1.59934617 1.12562869 0.29707354]
 [ 0.82159683 1.69283989 -0.40924636]
 [-1.25214437 -0.5302158 1.59119141]
 [-1.11823437 0.43625551 -1.31425178]
 [-1.0256384 -0.91681266 -1.42294103]]
```

Hasil Min-Max Scaling:

```
[[0.14557287 0.16982218 0.34800762]
 [0.74410555 0.2706176 0.20732587]
 [0.83455184 0.87317606 0.5399801 ]
 [0.50311809 0.09263167 0.9366178 ]
 [0.31020029 0. 0.74594236]
 [1. 0.80155398 0.57064997]
 [0.72724814 1. 0.33631391]
 [0. 0.22223583 1. ]
 [0.0469614 0.56036805 0.03605988]
 [0.07943423 0.08698003 0. ]]
```

IV. Kesimpulan

Normalisasi adalah metode untuk menghasilkan hasil yang sesuai dalam mencari normalisasi data. Teknik-teknik ini dapat digunakan untuk meningkatkan keterbacaan data dan menyederhanakan interpretasi, menyelesaikan masalah terkait perbedaan skala antara fitur, dan memprosesnya menjadi model pembelajaran mesin. Standardisasi dan penskalaan min-maks merupakan opsi yang harus dipertimbangkan bergantung pada persyaratan spesifik kumpulan data dan algoritma yang digunakan. Pengkodean ini merupakan implementasi dari dua fungsi untuk melakukan preprocess data pada suatu matriks menggunakan NumPy. Fungsi `np.random.seed(66)` untuk menentukan seed yang akan digunakan oleh generator nomor acak NumPy. `Standardize()` untuk menormalkan kumpulan data yang dihasilkan. `Min_max_scale()` (Variabel `scaled_data` berisi hasil) untuk menskalakan kumpulan data acak yang dinormalisasi dan diskalakan sesuai dengan prinsip min-max. Program di atas memenuhi tujuan dalam memperkenalkan konsep normalisasi data, mengaitkan konsep normalisasi data dengan kehidupan sehari-hari, Menguraikan tahapan normalisasi data.

REFERENCES

- [1] J.-M. Jo, "Effectiveness of Normalization Pre-Processing of Big Data," *The Journal of the Korea institute of electronic communication sciences*, vol. 14, no. 3, p. 548, 2019.
- [2] I. M. Karo Karo and Hendriyana, "KLASIFIKASI PENDERITA DIABETES MENGGUNAKAN ALGORITMA MACHINE LEARNINGDAN Z-SCORE," *Jurnaln Teknologi Terpadu*, vol. 8, no. 2, p. 96, 2022.
- [3] Henderi, T. Wahyuningsih and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *International Journal of Informatics and Information System*, vol. 4, no. 1, p. 16, 2021.
- [4] M. Shantal, Z. Othman and A. A. Bakar, "A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization," vol. 15, no. 12, p. 4, 2023.