

# **“Implementasi Python Closure Fuature engineering pada wine sales”**

## **Kelompok 7**

<sup>1</sup>Aditya Rahman  
122450113

<sup>2</sup>Farahanum Afifah Ardiansyah  
122450056

<sup>3</sup>Fayyaza Aqila Syafitri Achjar  
122450131

<sup>4</sup>Azizah Kusumah Putri  
122450068

<sup>5</sup>Eggi Satria  
122450032

<sup>6</sup>Meira Listyaningrum  
122450011

*Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera*  
*Jl. Terusan Ryacudu, Way Huwi, Kec. Jati Agung, Kabupaten Lampung Selatan, Lampung 35365*  
Kontak : <sup>1</sup>[aditya.122450113@student.itera.ac.id](mailto:aditya.122450113@student.itera.ac.id), <sup>2</sup>[farahanum.122450056@student.itera.ac.id](mailto:farahanum.122450056@student.itera.ac.id),  
<sup>3</sup>[fayyaza.122450131@student.itera.ac.id](mailto:fayyaza.122450131@student.itera.ac.id), <sup>4</sup>[azizah.122450068@student.itera.ac.id](mailto:azizah.122450068@student.itera.ac.id),  
<sup>5</sup>[eggi.122450032@student.itera.ac.id](mailto:eggi.122450032@student.itera.ac.id), <sup>6</sup>[meira.122450011@student.itera.ac.id](mailto:meira.122450011@student.itera.ac.id)

## **1. PENDAHULUAN**

### **1.1 Latar Belakang**

Pada saat ini, dengan perkembangan era digital yang berlangsung dengan pesat, jumlah data yang dihasilkan terus meningkat secara eksponensial. Teknik feature engineering merupakan bagian penting dalam pengembangan model prediktif untuk menciptakan representasi data yang sesuai.

Feature engineering adalah tahap penting dalam pengembangan model prediktif, yang melibatkan transformasi variabel-variabel data menjadi bentuk yang lebih representatif untuk meningkatkan kinerja model (Herdian et al., 2024, 11). Dengan menggunakan Feature engineering, para peneliti dan praktisi dapat mengubah, memanipulasi, dan mengekstrak fitur-fitur yang relevan dari data mentah, sehingga meningkatkan kemampuan model untuk mengidentifikasi pola dan membuat prediksi yang akurat.

Pada pembahasan kali ini, akan membahas bagaimana implementasi Python Closure pada feature engineer untuk melakukan analisis deskriptif seperti standarisasi dan normalisasi pada dataset penjualan wine.

### **1.2 Rumusan Masalah**

- Bagaimana Python Closure dapat diimplementasikan dalam feature engineer?
- Bagaimana feature engineer dapat diterapkan untuk melakukan standarisasi terhadap dataset?
- Bagaimana feature engineer dapat diterapkan untuk melakukan normalisasi terhadap dataset?

### **1.3 Tujuan**

Mengimplementasikan Python Closure dalam feature engineer untuk melakukan standarisasi dan normalisasi terhadap dataset penjualan wine.

## 2. METODE

### 2.1 Feature Engineering

Feature Engineering adalah suatu proses pengembangan dan pemilihan fitur atau atribut yang akan digunakan untuk analisis data dalam pembuatan model untuk machine learning (My Skill, 2023). Adapun tujuan dalam penggunaan feature engineering ini adalah untuk memperkaya representasi dataset dengan variabel-variabel penjelas tambahan dengan harapan agar dapat dihasilkan model prediksi yang lebih akurat dari informasi-informasi yang didapat dari variabel-variabel baru tersebut (Cerqueira et al., 2021).

### 2.2 Standarisasi

Standarisasi merupakan penentu suatu ukuran dalam produksi, dsb. Dalam statistika, standarisasi adalah salah satu teknik pengubahan skala dalam data sehingga memiliki rata-rata = 0 dan standar deviasi = 1 (Anzihory, 2021). Adapun formula yang digunakan untuk melakukan standarisasi terbagi menjadi dua berdasarkan asal datanya, yaitu data sampel dan data populasi

Formula standarisasi data sampel :

$$x_{standar} = \frac{x - \bar{x}}{S}$$

Formula standarisasi data populasi :

$$x_{standar} = \frac{x - \mu}{\sigma}$$

### 2.3 Normalisasi

Normalisasi dalam statistika merupakan teknik persiapan data dengan melakukan pengubahan pada kolom numerik menjadi skala yang lebih umum dalam suatu data agar semua data memiliki skala yang sama sehingga data dapat dibandingkan dengan lebih baik dan efektif (Trivusi, 2022). Normalisasi hanya memiliki satu formula untuk semua jenis data, tidak seperti standarisasi.

Formula normalisasi :

$$x_{normal} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

## 3. PEMBAHASAN

Dalam bab ini, kami melakukan analisis terhadap dataset penjualan wine yang mencakup data penjualan wine tahunan dengan berbagai atribut, termasuk tahun dan angka penjualan yang sesuai. Kami melakukan analisis deskriptif untuk memahami karakteristik data, termasuk rata-rata, median, dan distribusi data secara keseluruhan.

### 3.1 Menghitung Standarisasi

Teknik standarisasi memungkinkan kita agar bisa untuk mentransformasi data sehingga memiliki rata-rata nol dan deviasi standar satu. Dengan menggunakan closure, proses standarisasi dilakukan dengan langkah-langkah berikut:

- Hitung rata-rata( $\mu$ ) dan deviasi standar( $\sigma$ ) dari dataset
- Tentukan fungsi closure untuk dilakukannya standarisasi, yang akan mengambil data sebagai input dan mengembalikan data yang telah distandarisasi menggunakan rumus

$$x_{standar} = \frac{x - \mu}{\sigma}$$

Keterangan :

$x$  adalah nilai data

$\mu$  adalah rata-rata dari seluruh data

$\sigma$  adalah deviasi standar dari seluruh data

### 3.2 Menghitung Normalisasi

Teknik normalisasi memungkinkan kita untuk menskalakan data ke rentang antara 0 dan 1. Dengan menggunakan closure, proses normalisasi dilakukan dengan langkah-langkah berikut :

- Tentukan nilai minimum dan nilai maksimum dari  $x$  pada dataset
- Tentukan fungsi closure untuk normalisasi, yang akan mengambil data sebagai input dan mengembalikan data yang telah dinormalisasi menggunakan rumus

$$x_{normal} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Keterangan :

$x$  adalah nilai data

$\min(x)$  adalah nilai minimum dari seluruh data

$\max(x)$  adalah nilai maksimum dari seluruh data

Dengan menggunakan dan menerapkan metode ini kita dapat mengubah skala atau rentang data sehingga sesuai dengan kebutuhan analisis yang kita perlukan dan dapat menentukan perbandingan dan interpretasi yang lebih akurat

### 3.3 Analisis Hasil

Pada bagian ini kami akan melakukan analisis dari penerapan *closure* dalam *feature engineer* pada dataset penjualan wine. Pada pemrograman kami melakukan dua buah operasi, yaitu standarisasi dan

normalisasi. Standarisasi dilakukan dengan mengurangi setiap nilai dalam dataset dengan mean dari dataset, lalu hasilnya akan dibagi dengan standar deviasi dari dataset tersebut. Normalisasi dilakukan dengan mengurangi setiap nilai dalam dataset dengan nilai minimum dataset tersebut, lalu hasilnya akan dibagi dengan selisih nilai maksimum dan minimum dari dataset. Ini merupakan langkah-langkah yang cukup penting dalam melakukan analisis data karena akan membantu dan memastikan konsistensi serta keseragaman dalam dataset yang digunakan, yang pada akhirnya akan membantu mempermudah interpretasi dan analisis.

Kami membuat sebuah fungsi *closure* bernama *feature\_engineering\_tranfrom* untuk melakukan operasi *feature engineer* pada dataset yang diberikan. Fungsi ini akan menerima satu parameter yaitu *operation* yang merupakan fungsi lain yang akan diterapkan pada setiap elemen dalam dataset. Dengan menggunakan fungsi *closure* ini, kami membuat dua fungsi lain bernama '*standardize*' dan '*normalize*'. Kedua fungsi ini menggunakan fungsi *closure feature\_engineering\_tranfrom* untuk menerapkan operasi sesuai pada dataset. Setelah mendefinisikan fungsi-fungsi tersebut, kami menerapkannya dalam dataset, output akan menampilkan tabel yang telah ditambahkan dua kolom baru yaitu *winesales\_standardized* yang berisi data yang telah distandarisasi dan *winesales\_normalized* berisi data yang telah dinormalisasi.

Dalam tabel output, kita dapat mengamati bahwa dataset penjualan wine telah melalui proses *feature engineering* dibantu dengan fungsi *closure*. Tabel menampilkan beberapa baris data yang telah diproses, dengan setiap baris mewakili penjualan wine pada tanggal tertentu. Hasil dari pemrograman yang kami lakukan menunjukkan bahwa teknik *closure* telah berhasil diterapkan untuk melakukan operasi *feature engineer* pada dataset penjualan wine. Pendekatan ini memungkinkan kami untuk melakukan transformasi data seperti standarisasi dan normalisasi dengan menggunakan fungsi yang dapat digunakan kembali, ini meningkatkan efisiensi dalam proses analisis data.

Berikut link pemrograman dari penerapan *closure* dalam *feature engineer* menggunakan dataset penjualan wine dari tahun ketahun yang telah kami buat.

[https://colab.research.google.com/drive/1gX0-CrmZyeU\\_spFfnx6iNmkWUPM7ewym?usp=sharing](https://colab.research.google.com/drive/1gX0-CrmZyeU_spFfnx6iNmkWUPM7ewym?usp=sharing)

#### 4. KESIMPULAN

Berdasarkan dari pembahasan yang ada, dapat disimpulkan bahwa dalam *feature engineer*, *closure* dapat digunakan untuk membuat fungsi yang dapat dipersonalisasi dan memiliki akses ke parameter yang telah didefinisikan. *Feature engineering* dapat diterapkan untuk melakukan standarisasi terhadap dataset, dengan cara dengan mengurangi setiap nilai dalam dataset dengan mean dari dataset, lalu hasilnya akan dibagi dengan standar deviasi dari dataset tersebut. Selain menggunakan standarisasi, *feature engineering* juga dapat diterapkan untuk melakukan normalisasi terhadap dataset, dengan cara mengurangi setiap nilai dalam dataset dengan nilai minimum dataset tersebut, lalu hasilnya akan dibagi dengan selisih nilai maksimum dan minimum dari dataset.

#### References

- Anzihory, E. (2021, July 16). *Normalisasi vs Standarisasi*. Medium. Retrieved April 11, 2024, from <https://anzihory.medium.com/normalisasi-vs-standarisasi-101093633e18>
- Cerqueira, V., Moniz, N., & Soares, C. (2021, 4 6). VEST: automatic feature engineering for forecasting (T. Verdonck, M. Oskarsdottir, & S. v. Broucke, Eds.). *Springer*, (Machine Learning). <https://doi.org/10.1007/s10994-021-05959-y>
- Herdian, C., Kamila, A., Tampinongkol, F. F., Kembau, A. S., & Budidarma, I. G. A. M. (2024). One-hot encoding feature engineering untuk label-based data studi kasus prediksi harga mobil bekas. *Informasi Interaktif. Jurnal Informatika dan Teknologi Informasi*, 9(1), 10-16. <http://informasiinteraktif.janabadra.ac.id/index.php/jii/article/view/41>
- My Skill. (2023, September 25). *Feature Engineering*. Medium. Retrieved April 11, 2024, from <https://medium.com/@myskill.id/feature-engineering-327d64277211>
- Trivusi. (2022, September 16). *Normalisasi Data: Pengertian, Tujuan, dan Metodenya*. Trivusi. Retrieved April 11, 2024, from <https://www.trivusi.web.id/2022/09/normalisasi-data.html>