



Institut Informatika & Bisnis  
**DARMAJAYA**  
Keunggulan. Afiliasi. Mutu.



**Kampus  
Merdeka**  
INDONESIA JAYA

**MERDEKA  
BELAJAR**

# Statistika untuk Sains Data

## SSD23401

### PERTEMUAN 2

Egi Safitri, S.Mat., M.Si

`egisafitri@ darmajaya.ac.id`

October 10th, 2023



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan  
Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara  
Atribut dan Penggunaan  
Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan  
Mengklasifikasikan Data  
Metode Klustering dan  
Klasifikasi

### 1 Transformasi Data

- Transformasi Data
- Contoh Transformasi Data
- Alat dan Perpustakaan

### 2 Analisis Multivariat

- Analisis Multivariat dan Pola Tersembunyi
- Klasifikasi
- Pola Tersembunyi

### 3 Korelasi dan Regresi dalam EDA

- Analisis Korelasi antara Atribut dan Penggunaan Regresi
- Korelasi antara Atribut
- Regresi dalam EDA

### 4 Klustering dan Klasifikasi dalam EDA

- Mengelompokkan dan Mengklasifikasikan Data
- Metode Klustering
- Klasifikasi - Pengenalan





## Transformasi Data

### Transformasi Data

Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi



## Definition

Transformasi Data merujuk pada proses mengubah data mentah menjadi format atau bentuk lain yang lebih mudah diinterpretasikan atau diolah. Tujuannya adalah untuk mempermudah analisis data dan mengungkap informasi yang tersembunyi dalam dataset.

## Tujuan Transformasi Data

Tujuan utama dari Data Transformation adalah untuk mengatasi masalah dalam dataset, seperti *outlier* (nilai ekstrem), kekurangan data, atau ketidakcocokan format data. Selain itu, Transformasi data juga digunakan untuk meningkatkan kualitas data dan membuatnya lebih cocok untuk analisis statistik atau pemodelan.



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi



Berikut ini adalah beberapa contoh transformasi data yang umum digunakan dalam Eksplorasi Data (Exploratory Data Analysis - EDA) dan analisis statistik.

- Transformasi **Logaritmik**: Mengurangi skewness dan varians, cocok untuk data dengan hubungan multiplikatif seperti pendapatan atau populasi.
- Transformasi **Akar Kuadrat**: Mengurangi skewness dan varians, sesuai untuk data dengan batas bawah positif seperti hitungan atau luas.
- Transformasi **Berpangkat**: Mengurangi skewness dan varians dengan parameter lambda, ideal untuk data dengan berbagai nilai seperti tinggi atau berat.



## Transformasi Data

Transformasi Data

Contoh Transformasi Data

Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan

Pola Tersembunyi

Klasifikasi

Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi

Korelasi antara Atribut

Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data

Metode Klustering

Klasifikasi



- **Standarisasi:** Mengubah skala data sehingga memiliki rata-rata nol dan deviasi standar satu, berguna untuk variabel dengan skala atau unit yang berbeda seperti usia atau pendapatan.
- **Normalisasi Min-Max:** Mengubah skala data sehingga memiliki minimum nol dan maksimum satu, cocok untuk variabel dengan rentang tetap seperti persentase atau peringkat.



## Transformasi Data

Transformasi Data

Contoh Transformasi Data

Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi

Klasifikasi

Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi

Korelasi antara Atribut

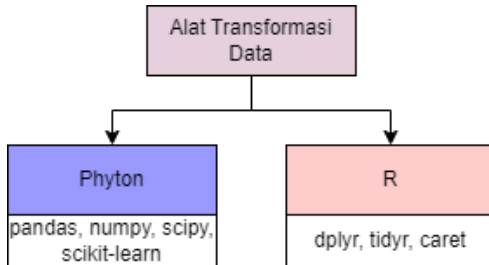
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data

Metode Klustering dan

Klasifikasi



Ada berbagai alat dan perpustakaan yang dapat digunakan dalam bahasa pemrograman Python dan R untuk menerapkan transformasi data sesuai kebutuhan analisis.



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi



## Definition

Analisis multivariat (multivariate analysis) merupakan salah satu jenis analisis statistik yang digunakan untuk menganalisis data yang terdiri dari banyak variabel baik variabel bebas (independent variables) maupun banyak variabel tak bebas (dependent variables).

Karena di dunia nyata, pengaruh berbagai faktor selalu bersifat kompleks dan simultan, dan analisis multivariat mampu untuk menjelaskan hal tersebut.



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

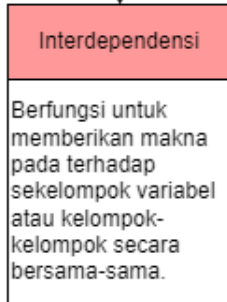
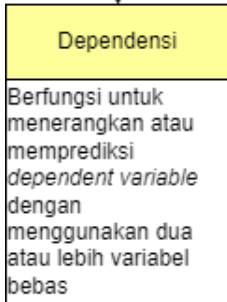
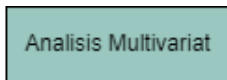
Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi





## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi

### 1 Dependensi:

- Adanya dua jenis variabel: variabel dependen dan variabel independen
- Antar variabel ada saling ketergantungan

### 2 Independensi:

- Semua variabel bersifat independen
- Variabel-variabel tidak saling bergantung satu dengan yang lain





## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi

Pola tersembunyi dalam analisis multivariat adalah informasi yang tidak terlihat saat kita hanya melihat satu hal dalam data. Ketika kita memeriksa beberapa hal sekaligus, kita bisa menemukan hubungan atau pola yang tidak akan kita temukan jika kita hanya memeriksa satu hal.

Misalnya, kita bisa menemukan hubungan yang kompleks antara beberapa hal, atau kita bisa mengelompokkan data menjadi kelompok-kelompok tertentu yang tidak terlihat jika kita hanya melihat satu hal. Dengan kata lain, analisis multivariat membantu kita melihat lebih dalam dan mendapatkan wawasan yang lebih baik tentang data kita.





## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi



Dalam eksplorasi data (EDA), kita sering melakukan analisis korelasi untuk mengukur hubungan antara atribut, dan juga menggunakan regresi untuk memahami dan memodelkan hubungan antara atribut dengan tujuan prediksi dan pemahaman lebih lanjut tentang data.



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi



Korelasi adalah ukuran statistik yang digunakan untuk mengukur sejauh mana dua variabel berkaitan satu sama lain. Ini membantu kita memahami apakah ada hubungan positif, negatif, atau tidak ada hubungan sama sekali antara dua atribut dalam data.

## Contoh

Korelasi antara suhu dan penjualan es krim di musim panas.



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi

- 1 Korelasi positif: Ketika peningkatan dalam satu atribut berhubungan dengan peningkatan dalam atribut lainnya.
- 2 Korelasi negatif: Ketika peningkatan dalam satu atribut berhubungan dengan penurunan dalam atribut lainnya.
- 3 Korelasi nol: Tidak ada hubungan yang jelas antara kedua atribut.





## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan  
Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara  
Atribut dan Penggunaan  
Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan  
Mengklasifikasikan Data  
Metode  
Klasifikasi



## Definition (Regresi)

Regresi adalah teknik statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (atribut) dengan variabel dependen (target). Ini memungkinkan kita untuk melakukan prediksi berdasarkan data yang ada.

## Contoh

Memprediksi harga rumah berdasarkan luas tanah, jumlah kamar, dll.



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klastering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klastering dan Klasifikasi



## Definition (Regresi linier)

Regresi linier adalah salah satu jenis regresi yang paling umum digunakan. Ini mengasumsikan hubungan linier antara variabel independen dan variabel dependen.

## Contoh

Regresi linear sederhana dan berganda.



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi



Klustering adalah teknik yang digunakan untuk mengelompokkan data serupa ke dalam kelompok-kelompok atau klaster. Ini membantu mengidentifikasi pola tersembunyi dalam data.

## Keuntungan Klustering

Klustering membantu kita memahami bagaimana data terstruktur dan membentuk kelompok-kelompok berdasarkan kesamaan karakteristik. Ini berguna untuk analisis pelanggan, segmentasi pasar, dan banyak aplikasi lainnya.



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data

Metode Klustering  
Klasifikasi



Terdapat berbagai metode klustering, yaitu :

- 1 K-Means
- 2 Hierarki
- 3 DBSCAN



## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klustering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klustering dan Klasifikasi

Klasifikasi adalah teknik yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas berdasarkan atribut-atribut tertentu. Ini membantu dalam membuat prediksi dan pengambilan keputusan.

## Keuntungan Klasifikasi Data

Klasifikasi membantu dalam memahami pola dan tren dalam data, memungkinkan pengklasifikasian data baru ke dalam kategori yang sesuai.





## Transformasi Data

Transformasi Data  
Contoh Transformasi Data  
Alat dan Perpustakaan

## Analisis Multivariat

Analisis Multivariat dan Pola Tersembunyi  
Klasifikasi  
Pola Tersembunyi

## Korelasi dan Regresi dalam EDA

Analisis Korelasi antara Atribut dan Penggunaan Regresi  
Korelasi antara Atribut  
Regresi dalam EDA

## Klastering dan Klasifikasi dalam EDA

Mengelompokkan dan Mengklasifikasikan Data  
Metode Klastering dan Klasifikasi



Terdapat berbagai metode klasifikasi, seperti :

- 1 Naive Bayes
- 2 Decision Trees
- 3 Support Vector Machines (SVM)

## Mengapa Klastering dan Klasifikasi penting dalam EDA

Klastering dan klasifikasi adalah alat penting dalam EDA yang membantu kita memahami dan mengelola data dengan lebih baik, mengungkap pola dan hubungan yang mungkin tersembunyi dalam data.

Thank You.

Egi Safitri, S.Mat., M.Si

