

## **BAB II**

### **LANDASAN TEORI**

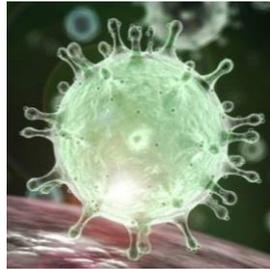
#### **2.1. COVID-19**

COVID-19 adalah penyakit yang disebabkan oleh virus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). COVID-19 dapat menyebabkan gangguan sistem pernapasan, mulai dari gejala yang ringan seperti flu, hingga infeksi paru-paru, seperti pneumonia. Kasus pertama penyakit ini terjadi di kota Wuhan, Cina, pada akhir Desember 2019 (dr. Pane, 2020).

Penyebaran virus Corona ini ditransmisikan melalui droplet (percikan air liur) yang dihasilkan saat orang yang terinfeksi batuk, bersin atau menghembuskan nafas. Droplet ini terlalu berat dan tidak bisa bertahan di udara, sehingga dengan cepat jatuh dan menempel pada lantai atau permukaan lainnya. Penularan dapat terjadi jika menyentuh permukaan benda yang terkontaminasi lalu menyentuh mata, hidung, atau mulut.

Gejala Covid-19 dapat bermacam-macam di antaranya demam tinggi lebih dari 38 derajat Celcius, batuk kering, lemas, sakit tenggorokan, sesak atau kesulitan bernapas, sakit kepala dan berbagai gejala yang lebih parah jika mengidap komplikasi penyakit yang sudah ada sebelumnya. Masa inkubasi virus ini sekitar 14 hari.

Gambar dari Virus Corona SARS-CoV-2 diambil dari pemindaian melalui mikroskop electron oleh *National Institute of Allergy and Infectious Diseases* (NIAID). Virus SARS-CoV-2 ini tidak berbeda jauh dengan struktur virus MERS-CoV dan SARS-CoV. struktur yang tidak berbeda jauh adalah pada bagian mahkota yang jelas terlihat.



Gambar 2.1. Bentuk Virus Corona  
sumber : <https://today.line.me/id/v2/article/B1ZZ6G>

## 2.2. Statistik

Statistika (Statistics) pada dasarnya berbeda dari Statistik (Statistic). Statistika merupakan ilmu yang berhubungan dengan pengumpulan, analisis, penginterpretasian data serta melakukan prediksi, sementara Statistik adalah data, informasi yang didapat dari data serta perhitungan dari suatu algoritma tertentu, tetapi sebutan Statistik sudah umum digunakan untuk semua yang berkaitan dengan Statistika maupun Statistik (Jimy, 2020).

Dua Jenis Statistik berdasarkan metodenya yaitu :

- Statistik Deskriptif / Non-Eksperimental

Statistik deskriptif berkaitan dengan deskripsi data, menggambarkan informasi dari suatu data tersebut misalnya rata-rata, median, modus (mode), standard deviasi dan varian dari sekumpulan data yang dapat dianalisa dan

divisualisasikan dengan tabel dan grafik agar mudah dibaca dan lebih bermakna.

- **Statistik Inferensial / Induktif / Probabilitas / Eksperimental**

Statistik Inferensial digunakan untuk melakukan pengujian hipotesis, melakukan prediksi di masa depan dengan regresi, atau membuat klasifikasi suatu data dengan cara membuat model dan biasanya digunakan untuk pengambilan keputusan berdasarkan analisis data.

### **2.3. Data**

Data merupakan elemen utama dalam statistik, data dapat dikelompokkan menjadi beberapa kategori tergantung dari jenis datanya (Jimy, 2020)

#### **1. Data Kualitatif / Kategorikal (Qualitative / Categorical)**

- **Nominal** yaitu data kategori yang tidak memiliki tingkatan dan setara satu sama lain misalnya jenis kelamin, warna rambut, wilayah, golongan darah.
- **Ordinal** yaitu data kategori yang memiliki tingkatan dan tidak setara satu sama lain dan memiliki ranking misalnya posisi dalam balap, kategori nilai A-F (A lebih tinggi nilainya dari F), Ya dan Tidak.

#### **2. Data Kuantitatif / Numerikal (Quantitative / Numerical)**

- **Diskrit / Discrete**, nilai diskrit berarti nilainya berupa bilangan bulat dan jumlahnya biasanya terbatas, misalnya jumlah orang (tidak bisa 0.5 orang), jumlah mobil, dan sebagainya.
- **Kontinu / Continuous**, nilai kontinu berarti nilainya bisa berupa bilangan pecahan dan jumlahnya tidak terbatas, misalnya jumlah uang, tinggi badan (bisa 167.53cm), berat badan, jarak dan sebagainya.

## 2.4. Analisis Regresi

Analisis regresi merupakan suatu analisis yang bertujuan untuk mengetahui pengaruh suatu variabel terhadap variabel lain. Model regresi paling sederhana adalah model regresi linier pada persamaan 1 (Hosmer, 2000), yaitu :

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.1)$$

Di mana :

$Y$  = variabel terikat (nilai yang diprediksi)

$X$  = variabel bebas

$\beta_0$  = konstanta

$\beta_1$  = koefisien regresi (nilai peningkatan atau penurunan)

$\varepsilon$  = galat acak

## 6.1. Regresi Linier

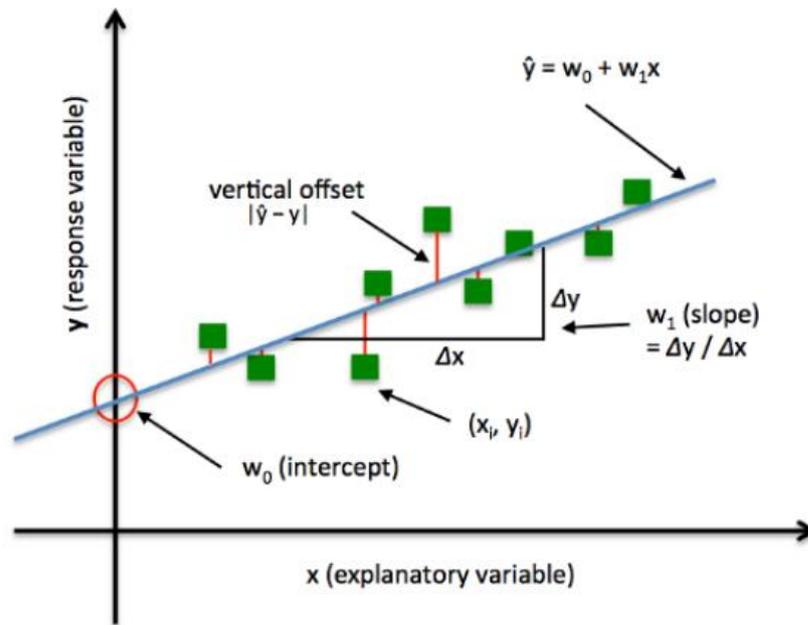
Regresi Linier atau Linear Regression adalah suatu model statistik yang umum dan paling sederhana yang digunakan untuk Machine Learning untuk melakukan prediksi dengan cara supervised Learning. Proses Regresi Linier adalah mencari garis linier yang menunjukkan korelasi antara kedua variabel tersebut, garis linier akan didapat dengan sebuah persamaan :

$$y = ax + b \quad (2.2)$$

Di mana :

$a$  adalah *slope / gradient / coefficient*

$b$  adalah *intercept*



Gambar 2.2. Data X dan Y yang tersebar dalam sumbu x dan y  
 sumber : <https://medium.com/@jrendz/regresi-linier-dengan-r-dan-python-ebb80662c6da>

### 2.5. Correlation Coefficient

Sebelum menggunakan regresi linier, korelasi antara variabel harus dihitung terlebih dahulu menggunakan Pearson Correlation ( $r$ ) untuk analisis bivariate :

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (2.3)$$

Di mana :

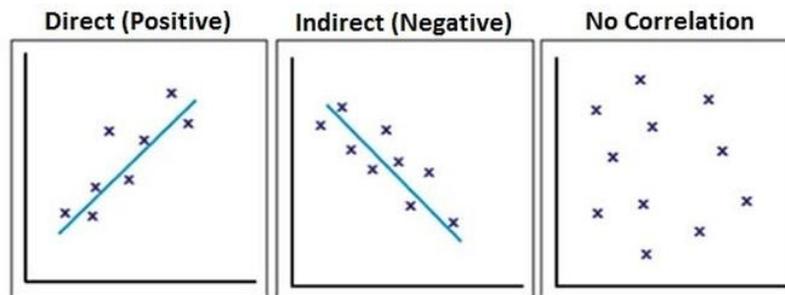
$x_i$  : jumlah hari

$\bar{x}$  : rata-rata (*mean*) jumlah hari

$y_i$  : jumlah kasus penderita pandemi virus SARS-CoV-2

$\bar{y}$  : rata-rata (*mean*) jumlah kasus penderita pandemic virus SARS-CoV-2

Jika Koefisien korelasi tinggi, maka titik-titik akan cenderung menurun atau naik. Rentang nilai koefisien korelasi adalah -1 sampai +1. Nilai 0 menunjukkan tidak ada korelasi antara kedua variabel. Jika nilai koefisien korelasi tinggi maka algoritma regresi linier cocok untuk data tersebut.



Gambar 2. 3. Koefisien Korelasi

sumber : <https://www.biologyforlife.com/correlation.html>

Setelah ditemukan korelasi tinggi, selanjutnya adalah mencari persamaan garis dengan dahulu nilai a (Slope) dan b (Intercept), dengan menggunakan rumus Least Square, rumus yang digunakan adalah sebagai berikut :

$$a = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad (2.4)$$

$$b = \bar{y} - a\bar{x} \quad (2.5)$$