

Logistic Regression Model

Logistic Regression Model adalah model statistik yang digunakan untuk memperkirakan probabilitas terjadinya suatu peristiwa berdasarkan dataset variabel bebas yang diberikan. Model ini sering digunakan untuk **klasifikasi** dan **analisis prediktif**. Hasil dari model ini berupa probabilitas yang berkisar antara **0** hingga **1**. Berikut adalah beberapa poin penting mengenai *Logistic Regression*:

Definisi: *Logistic Regression* (juga dikenal sebagai *logit model*) memodelkan log-odds dari suatu peristiwa sebagai kombinasi linear dari satu atau lebih variabel bebas.

Probabilitas: Model ini memperkirakan probabilitas terjadinya suatu peristiwa berdasarkan dataset variabel bebas. Karena hasilnya berupa probabilitas, variabel terikat dibatasi antara **0** dan **1**.

Transformasi Logit: Pada *Logistic Regression*, dilakukan transformasi logit pada odds—yaitu probabilitas keberhasilan dibagi dengan probabilitas kegagalan. Fungsi logistik ini direpresentasikan oleh rumus berikut:

$$\text{Logit}(\pi) = 1 / (1 + \exp(-\pi))$$

$$\ln(\pi / (1 - \pi)) = \text{Beta}_0 + \text{Beta}_1 X_1 + \dots + \text{Beta}_K X_K$$

Di sini, *logit* (π) adalah variabel terikat atau respons, dan x adalah variabel bebas. Parameter beta (koefisien) dalam model ini biasanya diestimasi melalui metode *maximum likelihood estimation* (MLE).

Interpretasi: Log-odds sulit untuk diinterpretasikan. Oleh karena itu, seringkali koefisien beta dieksponensiasi untuk mengubah hasil menjadi *odds ratio* (OR). OR menggambarkan peluang suatu hasil terjadi dibandingkan dengan peluang hasil tersebut tanpa adanya peristiwa tertentu. Jika OR lebih besar dari 1, maka peristiwa tersebut berhubungan dengan peluang lebih tinggi menghasilkan hasil tertentu. Sebaliknya, jika OR kurang dari 1, maka peristiwa tersebut berhubungan dengan peluang lebih rendah hasil tersebut terjadi. Contoh interpretasi OR: Jika kita memperkirakan odds bertahan hidup di kapal Titanic untuk penumpang pria, dan OR untuk pria adalah 0.0810, kita dapat menginterpretasikan OR tersebut sebagai peluang bertahan hidup penumpang pria berkurang sebesar faktor 0.0810 dibandingkan dengan penumpang wanita, dengan mempertahankan variabel lainnya konstan.