

# Regression Model

- ***Observation***

Data yang dikumpulkan sebagai informasi, secara umum mengacu pada 1 baris data yang terdiri dari beberapa variabel.

- ***Target variable***

Variabel yang ingin diprediksi/dimodelkan, sering disebut sebagai **respon/dependent variable**.

- ***Predictor***

Variabel yang digunakan untuk memprediksi **target variable**, sering disebut sebagai **independent variable**.

- ***Dummy variable***

Dummy Variable adalah hasil transformasi variabel kategorik dengan nilai 0 atau 1. Variabel ini digunakan untuk membuat data kategorik yang bersifat kualitatif menjadi kuantitatif.

- ***Outlier***

Data yang nilainya sangat ekstrim, sering disebut sebagai pencilan.

- ***Leverage***

Observation yang letaknya sangat berbeda (cukup jauh) dengan **observation** lain, dapat dikatakan sebagai data outlier.

- ***Influence***

Pengaruh yang ditimbulkan oleh **leverage** terhadap model regresi.

- ***Residual/Error***

Selisih antara nilai yang diprediksi dan nilai sebenarnya.

- ***Ordinary least square***

Memperkirakan parameter model regresi dengan meminimumkan nilai Sum Squared Error (SSE).

- ***Intercept***

Titik perpotongan antara sumbu y dengan garis regresi, merupakan salah satu **coefficient** regresi.

- ***Slope***

Kemiringan garis regresi, merupakan salah satu **coefficient** regresi.

- ***R Squared***

Persentase variansi **target variable** yang dapat dijelaskan oleh model (**predictor**). Ukuran yang bisa digunakan untuk mengukur kebaikan model.

- ***Adjusted R Squared***

Persentase variansi **target variable** yang dijelaskan oleh model, perbedaannya dengan **R Squared** adalah **Adjusted R squared** memperlakukan banyak observasi dan signifikansi prediktor yang digunakan.

- ***Akaike Information Criterion (AIC)***

Nilai yang menjelaskan besar informasi yang hilang pada model.

- ***Feature Engineering***

Feature engineering merupakan tahapan membuat **predictor** baru dari **predictor** yang sudah ada.

- ***Feature selection***

Feature selection merupakan tahapan dalam memilih variabel yang akan digunakan.

- ***Stepwise***

Stepwise adalah algoritma yang secara bertahap menambahkan atau mengurangi variabel dengan mengacu pada nilai AIC terkecil, diantaranya adalah **backward**, **forward**, **both**.

- ***Sum squared error (SSE)***

Jumlah dari **error** kuadrat. Ukuran yang bisa digunakan untuk mengukur **kebaikan model**.

- ***Mean squared error (MSE)***

Rata-rata dari **error** kuadrat. Ukuran yang bisa digunakan untuk mengukur **kebaikan model**.

- ***Root Mean squared error (RMSE)***

Akar kuadrat dari MSE.

- ***Normality***

Keadaan dimana residual yang dihasilkan berdistribusi normal.

- ***Linearity***

Keadaan dimana antara **predictor** dengan **target variable** memiliki hubungan yang **linear**.

- ***Homoscedasticity***

Residual yang dihasilkan bernilai konstan dan tidak membentuk pola apapun.

- ***Multicollinearity***

Keadaan dimana terdapat hubungan/**correlation** yang tinggi antar **predictor** pada model regresi.

- ***VIF (Variance Inflation Factor)***

Nilai yang dapat digunakan sebagai tolak ukur adanya **multicollinearity**.