

# Neural Network

- ***Target variable***

Variable yang nilainya dipengaruhi oleh **predictor** dan akan diprediksi nilainya, sering disebut sebagai respon/dependent variable.

- ***Predictor***

Variable yang mempengaruhi nilai **target variable** dan digunakan untuk memprediksi tersebut, sering disebut sebagai independent variable.

- ***Feature engineering***

Menambahkan informasi (variabel/kolom) berdasarkan informasi dari variabel lain yang sudah ada.

- ***Missing value***

Keadaan data memiliki nilai yang hilang (tidak diketahui nilainya).

- ***Standardization***

Proses untuk menyeragamkan skala data yang berbeda, sering disebut sebagai **scalling**.

- ***Data train***

Bagian data yang digunakan untuk membuat model (training model).

- ***Data test***

Bagian data yang digunakan untuk mengevaluasi performa model (testing model).

- ***Cross Validation***

Membagi data lengkap/utuh menjadi dua bagian data, yaitu **data train** dan **data test**.

- ***Class imbalance***

Keadaan dimana suatu kategori/level lebih mendominasi keseluruhan **target variable** (kelas mayoritas) dibandingkan kategori/level lainnya (kelas minoritas).

- ***Sampling***

Mengambil sebanyak  $n$  bagian data secara acak.

- ***Down-sample***

Proses **sampling** pada observasi kelas mayoritas, sebanyak jumlah observasi pada kelas minoritas. Tujuannya untuk menyamakan jumlah observasi pada kelas mayoritas dan minoritas.

- ***Up-sample***

Proses **sampling** pada observasi kelas minoritas, sebanyak jumlah observasi pada kelas mayoritas. Tujuannya untuk menyamakan jumlah observasi pada kelas mayoritas dan minoritas.

- ***Optimization***

Proses mengoptimalkan suatu nilai dengan menggunakan fungsi turunan, pada model neural network meminimumkan nilai error/kesalahan.

- ***Nodes***

Unit terkecil pada arsitektur neural network yang berfungsi untuk mengekstrak informasi (feature extraction) dan meneruskan informasi tersebut, sering disebut sebagai **neuron**.

- ***Input layer***

Lapisan pertama pada arsitektur neural network yang terdiri dari kumpulan **nodes**. Jumlah **nodes** pada **input layer** bergantung pada jumlah **predictor** pada data.

- ***Output layer***

Lapisan terakhir pada arsitektur neural network yang terdiri dari sebuah **nodes** atau beberapa **nodes** bergantung pada jenis **target variable**.

- ***Hidden layer***

Lapisan yang terletak di antara **input layer** dan **output layer**. Jumlah **hidden layer** dan jumlah **nodes** di setiap **hidden layer** ditentukan oleh peneliti.

- ***Weight***

Besar bobot yang menggambarkan besar informasi yang diteruskan dari setiap **nodes**. **Weight** ditetapkan secara acak.

- ***Linear regression***

Salah satu metode machine learning yang digunakan untuk memprediksi **target variable** bertipe numerik/angka.

- ***Bias***

Pada **linear regression** sama seperti nilai **intercept** ( $b_0$ ).

- ***Activation function***

Fungsi yang digunakan untuk mengubah interval nilai informasi yang masuk ke setiap **nodes** pada **hidden layer** dan **output layer**.

- ***Cost function***

Fungsi error yang digunakan pada model neural network.

- ***Feedforward***

Proses pada neural network yang dimulai dari **input layer** hingga menghasilkan nilai prediksi pada **output layer**.

- ***Backpropagation***

Proses pada neural network ketika melakukan **optimization** dan melakukan **update weight**.

- ***Epoch***

Satu kali proses **feedforward** dan **backpropagation**.

- ***Gradient***

Hasil turunan dari fungsi error/**cost function**.

- ***Dummy Variable***

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

- ***Learning rate***

Besar nilai yang menentukan seberapa cepat proses **update weight** hingga diperoleh nilai error yang konstan.

- ***Batch size***

Jumlah observasi yang diikutsertakan untuk satu iterasi.

- ***Confusion Matrix***

Metriks yang digunakan untuk mengukur kebaikan model classification, terdiri dari **accuracy**, **recall**, **specificity**, dan **precision**.

- ***ROC (Receiver Operating Characteristic)***

Kurva yang menggambarkan performa model klasifikasi untuk seluruh **threshold**.

- ***AUC***

Luas area di bawah kurva ROC, menggambarkan keberhasilan model klasifikasi dalam memprediksi/membedakan kedua kelas dari **target variable**.

- ***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.

- ***Mean absolute error (MAE)***

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