

Neural Network

- **Target variable**
Variable yang ingin diprediksi/dimodelkan, sering disebut sebagai **respon/dependent variable**.
- **Predictor**
Variable yang digunakan untuk memprediksi **target variable**, sering disebut sebagai **independent variable**.
- **Feature engineering**
Tahapan untuk menambah jumlah **variable**/kolom berdasarkan informasi dari **variable** lain yang sudah ada.
- **Missing value**
Keadaan dimana data memiliki nilai yang hilang (tidak diketahui nilainya).
- **Standardization**
Proses untuk menyeragamkan skala data yang berbeda, sering disebut sebagai **scaling**.
- **Data train**
Bagian data yang digunakan untuk membuat model.
- **Data test**
Bagian data yang digunakan untuk mengevaluasi kebaikan model.
- **Cross Validation**
Proses untuk membagi data menjadi dua bagian, 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**
Metode yang digunakan untuk meminimumkan error/kesalahan pada model neural network.
- **Node**
Unit terkecil pada arsitektur neural network yang berfungsi untuk melakukan transfer informasi, sering disebut sebagai **neuron**.
- **Input layer**
Lapisan pertama pada arsitektur neural network. Jumlah **node** pada **input layer** bergantung pada jumlah **predictor**.

- ***Output layer***

Lapisan terakhir pada arsitektur neural network. Jumlah **node** pada **output layer** bergantung pada jenis **target variable**.

- ***Hidden layer***

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

- ***Weight***

Besar bobot yang menggambarkan besar informasi yang diteruskan dari setiap **node**. **Weight** ditetapkan secara random (acak).

- ***Linear regression***

Suatu metode yang digunakan untuk memprediksi **target variable** bertipe numerik/angka.

- ***Bias***

Pada **linear regression** sama seperti nilai **intersept** (b_0). **Bias** ditetapkan secara random (acak).

- ***Activation function***

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

- ***Cost function***

Fungsi error.

- ***Feedforward***

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

- ***Backpropagation***

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

- ***Epoch***

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

- ***Gradient***

Hasil turunan dari **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 **gradient** yang digunakan untuk melakukan **update weight**.

- ***Batch size***

Jumlah observasi yang diikutsertakan untuk satu iterasi.

- ***Confusion Matrix***

Metrics 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.