**Glosarium**

Berikut adalah glosarium dengan istilah umum yang digunakan pada kelas ini. Anda dapat membaca sekilas materi berikut untuk mengenali istilah-istilah umum yang ada di modul kelas ini. Selain itu, Anda juga dapat mengunjungi kembali halaman ini setiap kali menemukan istilah yang belum dimengerti. Carilah istilah tersebut pada halaman glosarium ini untuk mengidentifikasi makna atau definisinya. Jika masih terdapat kosakata yang tidak Anda pahami dan belum masuk di daftar ini, Anda dapat memberikan saran melalui fitur Laporan Materi.

**A**

**Action**

Dalam reinforcement learning, action adalah setiap keputusan yang diambil oleh agent.

**Activation Function**

Fungsi yang menerima jumlah bobot semua masukan dari layer sebelumnya, kemudian menghasilkan dan meneruskan nilai keluaran ke layer berikutnya. Contoh: ReLU atau sigmoid.

**Akurasi**

Jumlah data yang diprediksi dengan benar oleh machine learning dibagi jumlah seluruh data poin.

**Algoritma**

Dalam ilmu komputer, algoritma adalah sekumpulan aturan atau instruksi yang didesain untuk melakukan tugas dan menyelesaikan permasalahan.

**Anomaly Detection**

Proses untuk menemukan anomali pada dataset.

**Artificial Intelligence**

Program atau model yang dapat menyelesaikan tugas-tugas canggih. Machine learning adalah sub-bidang dari artificial intelligence.

**B**

**Bobot**

Koefisien untuk fitur dalam model linier atau parameter yang dipelajari oleh sebuah perceptron untuk menunjukkan kekuatan node tertentu dalam jaringan syaraf tiruan.

**C**

**Clustering**

Pengelompokan data yang memiliki kesamaan ke dalam grup tertentu.

**CSV**

Comma-Separated Value, format file teks dalam bentuk tabular dengan pemisah berupa koma

**D**

**Data Kategorik**

Fitur-fitur yang memiliki sekumpulan nilai diskrit dan bisa dibagi ke dalam grup. Sering disebut juga data diskrit.

**Data Numerik**

Fitur-fitur yang direpresentasikan sebagai bilangan bulat atau bilangan riil. Sering disebut juga fitur berkelanjutan.

**Data Time Series**

Sekumpulan data atau observasi pada interval waktu tertentu.

**Dataset**

Sekumpulan data atau contoh-contoh yang terdiri dari satu atau lebih fitur.

**Deep Learning**

Cabang machine learning dengan algoritma jaringan syaraf tiruan yang dapat belajar dan beradaptasi terhadap sejumlah besar data.

**Dimension Reduction**

Mengurangi jumlah dimensi yang digunakan untuk merepresentasikan fitur tertentu.

**E**

**Elbow**

Metode untuk menentukan jumlah cluster pada dataset.

**Epoch**

Satu proses pelatihan penuh atas seluruh dataset sehingga setiap contoh data telah melewati model sebanyak satu kali.

**Environment**

Dalam reinforcement learning, environment adalah sarana untuk berinteraksi, yang dapat menerima action dan memberikan respon berupa hasil maupun data berupa satu set observasi baru.

**F**

**Feedback**

Dalam machine learning, situasi di mana prediksi model mempengaruhi data pelatihan.

**Fitur**

Input variabel yang digunakan untuk membuat prediksi.

**G**

**Good Fit**

Kondisi saat model model machine learning melakukan generalisasi data dengan baik.

**Google Colaboratory**

Notebooks yang memungkinkan kita menulis dan mengeksekusi code pada browser.

**H**

**Hidden Layer**

Lapisan sintetis dalam jaringan saraf antara lapisan masukan (fitur) dan lapisan keluaran (prediksi). Hidden layer biasanya berisi fungsi aktivasi untuk pelatihan.

**Hyperparameter**

Variabel yang digunakan untuk mengontrol proses pelatihan model. Contohnya: epoch.

**I**

**Input Layer**

Layer pertama yang menerima data masukan dalam jaringan saraf tiruan

**K**

**Keras**

API Python machine learning yang populer. Keras bisa dijalankan pada beberapa kerangka kerja deep learning, termasuk TensorFlow yang tersedia sebagai tf.keras.

**Klasifikasi**

Tipe model machine learning untuk membedakan antara dua atau lebih kelas. Misalnya klasifikasi apakah suatu email merupakan email spam atau bukan.

**L**

**Label**

Dalam supervised learning, label adalah “jawaban” atau “hasil” dari sebuah contoh.

**Layer**

Sekumpulan neuron dalam jaringan saraf yang memproses fitur-fitur masukan atau keluaran dari neuron tersebut.

**Linear Regression**

Menggunakan keluaran y’ dari model linier sebagai prediksi aktual dalam model regresi. Tujuan dari permasalahan regresi adalah untuk membuat prediksi yang bernilai benar.

**Logistic Regression**

Model klasifikasi yang menggunakan fungsi sigmoid untuk mengubah prediksi y’ pada model linier menjadi nilai antara 0 dan 1.

**M**

**Machine Learning**

Bidang studi yang memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit.

**Matplotlib**

Open-source library pada Python untuk membuat plot 2D. Matplotlib membantu Anda memvisualisasikan berbagai aspek dalam machine learning.

**Missing Value**

Tidak tersedia nilai data pada variabel tertentu dalam sebuah observasi.

**Model**

Representasi dari apa yang telah dipelajari oleh sistem machine learning dari data pelatihan. Model mendefinisikan relasi antara fitur dan label.

**N**

**Neural Network**

Sebuah model yang mengambil inspirasi dari otak, terdiri dari beberapa layers yang memiliki neuron-neuron yang saling terhubung.

**Normalization**

Proses mengonversi rentang nilai aktual menjadi rentang nilai standar, biasanya dari -1 hingga +1, atau 0 hingga 1.

**Numpy**

Library matematika open-source yang menyediakan operasi array effisien dengan Python. Pandas dibangun di atas NumPy.

**O**

**One-hot-encoding**

Mengubah data kategorik dengan membuat kolom baru untuk setiap kategori.

**Outlier**

Sebuah nilai yang jauh berbeda dari kumpulan nilai lainnya dan dapat mengacaukan hasil dari sebuah analisis statistik.

**Output layer**

Lapisan “terakhir” dari jaringan saraf tiruan. Lapisan yang berisi jawaban.

**Overfitting**

Kondisi model sangat cocok dengan data pelatihan, tetapi model gagal membuat prediksi yang benar pada data baru.

**P**

**Pandas**

API analisis data yang berorientasi pada kolom. Banyak framework machine learning termasuk TensorFlow mendukung struktur data panda sebagai input.

**Parameter**

Variabel model yang dilatih oleh sistem machine learning. Sebagai contoh: weight/bobot adalah parameter yang nilainya dipelajari secara bertahap oleh sistem machine learning.

**Propagasi balik**

Algoritma untuk melakukan penurunan gradien pada jaringan saraf tiruan. Pertama, nilai keluaran dari setiap node dihitung dalam sebuah forward pass. Kemudian, turunan parsial dari eror/kesalahan yang terkait dengan setiap parameter dihitung dalam hitungan mundur melalui grafik.

**R**

**Reinforcement Learning**

Algoritma yang belajar menggunakan sistem reward dan pinalti. Algoritma Reinforcement Learning belajar agar terus mendapatkan reward dan menghindari penalti.

**Reward**

Dalam reinforcement learning, reward adalah hasil numerik dari mengambil tindakan dalam suatu keadaan seperti yang didefinisikan oleh environment. Reward diberikan saat agent berhasil menyelesaikan tantangan.

**S**

**Scikit learn**

Platform machine learning open-source yang populer. Lihat [www.scikit-learn.org](https://scikit-learn.org/stable/).

**SVM**

Algoritma training yang bertujuan untuk memaksimalkan margin antara pola pelatihan dan batas keputusan (decision boundary).

**T**

**Tensorflow**

Platform machine learning berskala besar dan terdistribusi. Istilah ini juga merujuk pada layer API dasar yang mendukung komputasi umum pada grafik aliran data.

**Test set**

Bagian dari dataset yang digunakan untuk menguji sebuah model setelah model melalui pemeriksaan awal oleh validation set.

**Training set**

Bagian dari dataset yang digunakan untuk melatih model.

**U**

**Underfitting**

Model machine learning memiliki kemampuan prediksi yang buruk karena model tersebut belum menangkap kompleksitas data pelatihan.

**Unsupervised learning**

Model machine learning yang belajar dan menemukan pola dalam sekumpulan data tanpa label.

**V**

**Validation set**

Bagian dari dataset (berbeda dari training set) yang digunakan sebagai validasi.

Pada supervised learning kita melatih sekumpulan data yang memiliki label. Label adalah pengenal dari sebuah data. Misal buah A memiliki atribut atau fitur berwarna hijau, bentuknya bulat, dan ukurannya sebesar bola sepak. Buah B atributnya berwarna kuning, bentuknya agak memanjang, dan besarnya segenggaman tangan. Buah yang memiliki atribut seperti yang disebutkan pada buah A dikenali sebagai semangka, sedangkan buah dengan atribut seperti disebutkan pada buah B dikenal sebagai pisang. Maka label dari buah A adalah semangka dan label buah B adalah pisang.

