

1. Salah satu contoh case machine learning adalah mendeteksi penyakit.
2. Supervised learning dan unsupervised learning. Supervised learning memiliki target yang ingin dicapai sedangkan unsupervised learning tidak memiliki target seperti clustering.
3. Validation adalah set data untuk model mengecek kemampuan mereka predicting dan akan mendapatkan metric. Sedangkan testing adalah data set untuk dipelajari oleh model.
4. Classification digunakan untuk mengelompokkan target yang ingin ditujui sedangkan regression mencari garis sebagai pacuan untuk memprediksi target.
5. Overfitting akan terjadi bila model diajari dengan data yang sama berulang kali sedangkan underfitting karena model kurang mendapatkan data.
6. Cross-validation (CV) adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi / evaluasi. Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi. Selanjutnya pemilihan jenis CV dapat didasarkan pada ukuran dataset.
7. Discrete variable adalah variable yang didapatkan dari menghitung sedangkan continuous adalah data yang didapatkan dari mengukur.
8. Normal distribution data ini sering ditemukan dalam kehidupan nyata.
- 9.
10. Statistik parametrik adalah teknik statistik yang dapat digunakan untuk menguji hipotesis dengan melibatkan parameter populasi. Keterbatasan yang dimiliki oleh statistik parametrik adalah penggunaan jenis data minimal yaitu jenis data interval dan rasio. Lain halnya dengan statistik non parametrik, pengolahan data statistik ini tidak melakukan pendugaan sebelumnya.
11. Mean merupakan rata rata dari sebuah data sedangkan varians adalah rata-rata deviasi kuadrat dari mean.
12. Regularisasi adalah teknik yang digunakan untuk melakukan modifikasi pada model neural network yang bertujuan untuk mengurangi generalization error. Pada regularisasi digunakan untuk mengurangi keefektifan dari model itu.