Dendeni

1. Salah satu contoh case machine learning yang saya ketahui adalah case titanic. Dalam dataset tersebut, terdapat beberapa feature yang bisa digunakan, dimulai dari Parent/Child, PassengerId, Passenger Class, dan lain sebagainya. Kita juga diminta untuk memprediksi apakah passenger tersebut *Survived / Not Survived.* Kasus ini masuk dalam kasus yang dapat diprediksi dengan menggunakan machine learning.
2. Terdapat 3 tipe machine learning, pertama Supervised, Unsupervised, dan Reinforcement Learning. Tipe tipe ini umumnya dibedakan berdasarkan *past record data* yang dipunyai. Supervised sudah diberikan jawaban atau hasil target dari features, Unsupervised tidak diberikan hasil target sebelumnya, dan untuk reinforcement learning, terdapat suatu ‘agent’ yang akan ‘melatih’ model kita. Contoh supervised: regression, classification, dan lain sebagainya, Unsupervised: Clustering, dan reinforcement learning seperti *Autonomous Car*.
3. Validation merupakan data yang digunakan untuk melatih training set kita agar mendapat model yang prediktif. Testing set merupakan data yang digunakan untuk mengetest model prediktif kita terhadap data yang belum pernah dipakai. Pada dasarnya, dataset dibagi 80% training set dan 20% testing set, dalam 80% training set ini, akan dibagi lagi 80% menjadi training set dan 20% menjadi validation set
4. Cara kerja classification adalah dengan melakukan model model untuk mengklasifikasikan beberapa data menjadi beberapa bagian, misalnya survived / not survived. Model akan memprediksi apakah data seseorang survive atau tidak. Classification dapat dicek akurasinya dengan melihat precision, accuracy, metrics, atau f1 score dengan tujuan melihat berapa prediksi yang benar. Cara kerja regression adalah untuk memprediksi nilai yang tidak berkategori (continous) contohnya menebak harga suatu produk. Akurasi dari regression dapat dicek dengan menselisihkan nilai prediksi dengan nilai yang aktual
5. Training data bisa mengalami overfitting jika training data tersebut ‘menghafal’ pola dari data tersebut, dengan menghafal ini, model tersebut tidak dapat memprediksi data yang belum pernah dilihat model itu sebelumnya, hal ini bisa dilihat jika akurasi training tinggi sedangkan testing rendah. Selanjutnya, training data bisa mengalami underfitting jika data tersebut tidak cukup belajar dan menerima pola, ini bisa terjadi karena feature yang tidak dapat mendapatkan pola target data, pada kondisi ini, akurasi training dan test set rendah.
6. Cross Validation merupakan proses membagi pola splitting training dan testing set menjadi beberapa bagian. Misal training dan test set terbagi 80% dan 20% dan cv = 5, maka akan terdapat 5 pola berbeda dalam pembagian training dan testing set. Hal ini sering dilakukan untuk mengecek apakah model itu secara general sudah bagus atau belum.
7. Discrete random variable merupakan variable yang dimana nilainya tergantung pada rentang angka tertentu contohnya peluang kemunculan angka dari dadu, sedangkan continous random variable adalah variable yang tidak terbatas rentang angka, contohnya berat badan seseorang
8. – Uniform Distribution: merupakan pemodelan kejadian yang memilki kemungkinan sama untuk semua kejadian yang ada, contoh: pelemparan dadu

- Binomial distributionL merupakan pemodelan yang memungkinkan mendapatkan 1 dari 2 hasil yang ada, contoh: pelemparan koin

1. Mean merupakan nilai rata-rata suatu variable yang didapatkan dengan menjumlahkan seluruh angka dibagi banyaknya angka tersebut. Variance merupakan nilai rata-rata selisih variable dengan mean. Variance dapat dihitung dengan mencari mean, melakukan selisih mean terhadap tiap variable, lalu dijumlahkan, dikuadratkan, lalu dibagi dengan jumlah data yang ada.
2. Model parametrik memilki jumlah parameter yang ditetapkan sesuai dengan ukuran sampel yang ada, sednagkan model nonparametrik, jumlah parameter tumbuh dengan ukuran sampel. Contoh parameter dalam model parametrik adalah parameter pada regresi linear, berupa koefisien, dan non-parametric seperti jaringan neural yang hanya membuat data yang paling mirip dengan prediksi.
3. Loss function merupakan suatu bagian yang dapat menjadi alat pembelajaran bagi model kita agar ke depannya model kita dapat meminimalisir loss function baik dengan mengubah parameter pada tiap looping. Semakin kecil loss function, maka model semakin baik dalam memprediksi target variable
4. Regularization merupakan proses mengurangi hambatan pada proses meng-generalisasi model dengan mengurangi tinggi persebaran data dengan tetap mempertahankan bentuk model data. Optimization merupakan proses mengurangi ‘cost function’ model.