

# MACHINE LEARNING PREDIKSI

Arif Laksito, M. Kom www.ariflaksito.net



### MACHINE LEARNING

Machine learning secara umum terbagi menjadi 2 kategori, yaitu:

#### Supervised learning

Supervised learning adalah kategori machine learning yang menyertakan solusi yang diinginkan (label) dalam proses pembelajarannya.

Dataset yang digunakan telah memiliki label dan algoritma kemudian mempelajari pola dari pasangan data dan label tersebut.

### Unsupervised learning

Model unsupervised learning melakukan proses belajar sendiri untuk melabeli atau mengelompokkan data.

Dataset yang digunakan tidak memiliki label.



### ALGORITMA MACHINE LEARNING

Berikut Algoritma Machine learning berdasarkan kategorinya:

### Supervised learning

- Algoritma Prediksi: Linear regresion, Neural network, SVM, dll
- Evaluasi: RMSE, MSE, MAPE
- Algoritma Klasifikasi: Naïve bayes, KNN, Decission tree, Random forest, all
- Evaluasi: Confussion matrix, Accuracy

### Unsupervised learning

- Algoritma Klustering: K-Means, K-Medoids, Fuzzy C-Means
- Evaluasi: Davis-bouldin index, Dunn index



### PREDIKSI – REGRESI LINIER

- Suatu model statistik yang umum digunakan pada Machine learning untuk melakukan prediksi dengan cara supervised learning adalah Regresi linear.
- Regresi linier melibatkan 2 variabel dimana salah satunya adalah variabel independen (x) dan satu lagi adalah variabel dependen (y).
- Independen berarti variabel ini sebagai variabel utama yang mungkin akan mempengaruhi nilai variabel kedua (dependen).
- Dependen berarti nilai variabel ini akan tergantung dari nilai variabel independennya, jika korelasi tinggi maka dependensi juga tinggi.



### REGRESI LINIER

Regresi Linear adalah proses mencari garis linear yang menujukkan korelasi antara kedua variabel tersebut, garis linear itu akan didapat dengan sebuah persamaan:

$$y = ax + b atau y = w1x + w0$$

#### Dimana:

- a atau w1 adalah slope / gradient / coefficient
- b atau w0 adalah intercept / bias





Algoritma Regresi linier digunakan untuk Prediksi dengan mencari pola garis terbaik antara variable independent dan dependen

#### Pros:

- Mudah diimplementasikan
- Digunakan untuk memprediksi nilai numerik/continous

#### Cons:

- Cenderung mudah overfitting
- Tidak dapat digunakan bila relasi antara variabel independen dan dependen tidak linier atau korelasi variabel rendah



### **EVALUASI RMSE**

- Untuk mengetahui apakah model tersebut baik atau tidak, pada kasus regresi biasanya akan dilihat perbedaan output hasil prediksi dengan output sebenarnya.
- Misalkan dengan input 15, output hasil prediksi adalah 61, padahal output sebenarnya harusnya 59. Maka (61 - 59) = 2. Perbedaan tersebut dikuadratkan, 2^2 = 4.
- Angka tersebut kemudian dihitung rata-ratanya dengan perbedaan hasil prediksi lain, kemudian di-akar-kuadratkan. Metriks ini dinamakan Root Mean Square Error (RMSE), yang bisa ditulis seperti berikut:

RMSE = sqrt(average(error^2))



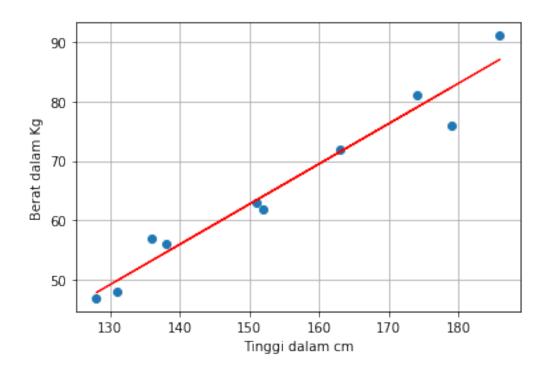


Terdapat data sample yang didapat dari 10 orang sebagai berikut :

Data Tinggi Badan(cm): 151,174,138,186,128,136,179,163,152,131

Data Berat Badan (Kg): 63, 81, 56, 91, 47, 57, 76, 72, 62, 48

Disini tinggi badan merupakan variabel independen (x), dan berat badan merupakan variabel dependen (y). Data x dan y ini dapat kita sebarkan ke dalam sumbu x dan y seperti berikut :







- Dicoding Belajar Machine learning untuk pemula, https://www.dicoding.com/academies/184/tutorials/8326
- Jimy Regresi Linier dengan R dan Python, https://medium.com/@jrendz/regresi-linier-dengan-r-dan-pythonebb80662c6da



## **TERIMAKASIH**

about us:

www.tlab.co.id