

MACHINE LEARNING KLASIFIKASI

Arif Laksito, M. Kom www.ariflaksito.net



KLASIFIKASI

- Klasifikasi adalah suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target.
- Klasifikasi dikategorisasikan ke dalam kategori supervised learning atau pembelajaran yang diawasi.
- Di Supervised learning kita memberi tahu kepada mesin (algoritma) cara pembagian kelompoknya, atau dengan kata lain kita memberikan knowledge ke mesin.
- > Algoritma Supervised learning yang akan kita gunakan:
 - 1. Naïve bayes
 - 2. Decision Tree
 - 3. K-Nearest neighbor (KNN)



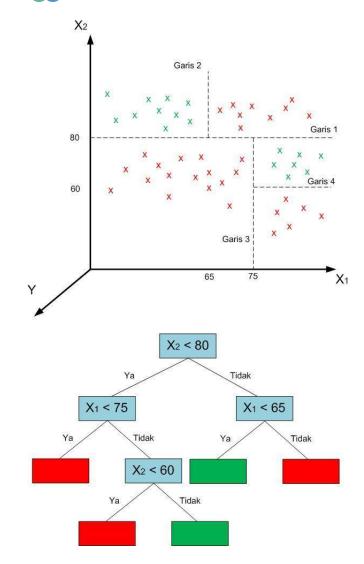
NAÏVE BAYES

- Algoritma Naive Bayes merupakan sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes.
- Keuntungan penggunan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter ya diperlukan dalam proses pengklasifikasian.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$



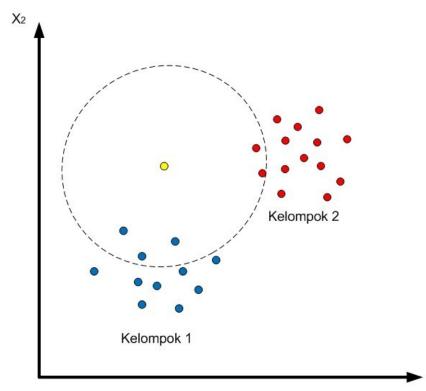
DECISION TREE



- Decision tree adalah salah satu metode klasifikasi yang paling populer, karena mudah untuk diinterpretasi oleh manusia. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.
- Konsep dari pohon keputusan adalah mengubah data menjadi decision tree dan aturan-aturan keputusan.
- Nama lain dari decision tree adalah CART (Classification and Regression Tree). Dimana metode ini merupakan gabungan dari dua jenis pohon, yaitu classification tree dan juga regression tree



K-NEAREST NEIGHBOR



- K-Nearest Neighbor (K-NN) adalah sebuah metode klasifikasi terhadap sekumpulan data yang diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN.
- Dekat atau jauhnya tetangga biasanya dihitung berdasarkan Euclidean distance, atau dapat juga menggunakan Manhattan distance atau Minkowski distance.



EVALUASI

- Untuk mengetahui apakah model tersebut baik atau tidak, pada kasus klasifikasi biasanya akan dilihat akurasinya, dengan cara menghitung jumlah prediksi benar dibagi jumlah total prediksi.
- Misalkan dari 50 prediksi, terdapat 3 prediksi yang salah (karena memprediksi sebagai laki-laki padahal harusnya perempuan, atau sebaliknya). Berarti akurasinya adalah 47/50 * 100% = 94%.

Accuracy = correct predictions / total predictions * 100

- Ciri dari problem klasifikasi adalah yang diprediksi adalah class, atau jenis.
- Outputnya bersifat discrete, dan biasanya terbatas (misal 2 class, 3 class, atau 10 class). Contoh lain dari klasifikasi ini misalnya klasifikasi jenis bunga iris, klasifikasi anjing atau kucing, klasifikasi spam email





- Kurnia Sari Pratiwi Naive Bayes Classifier using Scikit-learn in Python, https://medium.com/@kurniasp/naive-bayes-classifier-using-scikit-learn-inpython-3067144af115
- IYKRA Mengenal Decision Tree dan Manfaatnya, https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnyab98cf3cf6a8d
- MB Herlambang Machine Learning: Decision Tree Classification, https://www.megabagus.id/machine-learning-decision-tree-classification/
- MB Herlambang Machine Learning: K-nearest Neighbors, https://www.megabagus.id/machine-learning-k-nearest-neighbors/
- Informatikalogi Algoritma K-Nearest Neighbor (K-NN), https://informatikalogi.com/algoritma-k-nn-k-nearest-neighbor/



TERIMAKASIH

about us:

www.tlab.co.id