

TUGAS TEKNIK KLASIFIKASI

Mata Kuliah Data Science

Dosen Pengampu : Vynska Amalia Permadi S.Kom., M.Kom



Disusun oleh :

NAMA : Alivi Milova

NIM : 123170062

Kelas : C

PROGRAM STUDI INFORMATIKA

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK INDUSTRI

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

YOGYAKARTA

2020

TEKNIK KLASIFIKASI

Teknik Klasifikasi adalah training set yang berisi sekumpulan atribut, salah satu atributnya adalah kelas yang kemudian mencari model untuk atribut kelas sebagai fungsi dari nilai atribut lainnya dengan tujuan menentukan pembagian kelas yang akurat. Adapun jenis klasifikasi sebagai berikut :

1. Decision Tree based Methods

❖ Algoritma yang digunakan untuk Teknik klasifikasi Decision Tree ini adalah

a. Hunt's Algorithm

b. CART

c. ID3, C4.5

Ada beberapa hal yang perlu diperhatikan dalam penyelesaian masalah dengan menggunakan algoritma C4.5 antara lain

- Depth-first construction yang sederhana
- Menggunakan penguatan informasi
- Mengurutkan atribut berkelanjutan di setiap node.
- Membutuhkan seluruh data agar sesuai dengan memori.
- Tidak cocok untuk kumpulan data besar.
- Membutuhkan penyortiran out-of-core.

d. SLIQ, SPRINT

e. Nearest-Neighbor

Algoritma Nearest-Neighbor membutuhkan tiga hal yaitu

1. Set rekaman yang disimpan
2. Distance Metric untuk menghitung jarak antar rekaman
3. Nilai k, jumlah tetangga terdekat yang akan diambil

Untuk mengklasifikasikan catatan yang tidak dikenal:

1. Hitung jarak ke catatan pelatihan lainnya
2. Identifikasi k tetangga terdekat
3. Gunakan label kelas dari tetangga terdekat untuk menentukan label kelas dari catatan yang tidak diketahui (contoh : Dengan mengambil suara terbanyak).

❖ Cara Menentukan Test Condition pada Teknik Decision Tree

a. Tergantung pada tipe atribut

- Nominal
- Urut
- Kontinu

b. Tergantung pada beberapa cara untuk membagi

- 2-way split
- Multi-way split

❖ Keuntungan Decision Tree antara lain :

- a. Murah untuk dibangun
- b. Sangat cepat dalam mengklasifikasikan catatan yang tidak diketahui
- c. Mudah diinterpretasikan untuk pohon berukuran kecil
- d. Akurasi sebanding dengan teknik klasifikasi lainnya untuk banyak set data sederhana

2. Instance-based Methods

3. Memory based reasoning

4. Neural Networks

5. Naïve Bayes and Bayesian Belief Networks

Yaitu mempertimbangkan setiap atribut dan label kelas sebagai variabel acak

6. Support Vector Machines