

1. Deskripsi Masalah

Naïve Bayes Classifier adalah salah satu teknik dalam pembelajaran mesin yang digunakan untuk mengelompokkan data. Urutan algoritma naïve bayes adalah sebagai berikut :

- Menghitung jumlah kelas
- Menghitung jumlah kondisi per kelas
- Mengalikan semua hasil
- Membandingkan

1. Menghitung Jumlah kelas

Pertama kita perlu mengetahui jumlah kelas dari data yang kita miliki, jadi kita perlu membagi setiap kelas yang ada di dataTrain, pada kasus ini terdapat 1 output dengan 2 kelas, yaitu : income >50K dan income <=50K

2. Menghitung jumlah kondisi per kelas

Setelah mendapatkan jumlah kelas, kita perlu membandingkan setiap atribut terhadap kelas yang telah dibagi sebelumnya. Misalnya atribut Age dengan kelas young dengan output income dengan kelas >50K. Maka perhitungannya akan jadi seperti berikut:

$$A = \text{Income} > 50k / \text{SeluruhIncome}$$

$$\text{AGE1} = \text{AgeYoung} / A$$

3. Mengalikan semua hasil

Dimana setiap hasil yang didapatkan dari operasi yang dilakukan diatas akan dikalikan agar mendapatkan hasil untuk selanjutnya dibandingkan.

Kita akan mendapatkan 2 nilai, yaitu untuk kelas >50K dan <=50K, oleh karena itu kita perlu membandingkannya

4. Membandingkan

Hasil yang telah didapatkan dari proses sebelumnya akan dibandingkan. Hasil yang lebih besar adalah yang akan menempati kelas tersebut.

2. Analisis

Pada permasalahan kali ini terdapat 7 Atribut dimana masing masing atribut memiliki 2 sampai 3 kelas. Terdapat sejumlah 160 data latih yang digunakan untuk latihan agar bisa menyelesaikan klasifikasi 40 data tes.

160 data tersebut digunakan untuk mengetahui rataan dari masing masing atribut dengan masing masing kelasnya untuk klasifikasi data tes yang telah disediakan

Berikut adalah hasil runningnya :

```
-
0  <=50K
1  >50K
2  >50K
3  >50K
4  <=50K
5  >50K
6  <=50K
7  <=50K
8  >50K
9  >50K
10 <=50K
11 <=50K
12 <=50K
13 >50K
14 >50K
15 >50K
16 >50K
17 <=50K
18 >50K
19 >50K
20 >50K
21 >50K
22 <=50K
23 >50K
24 <=50K
25 >50K
26 >50K
27 >50K
28 >50K
29 >50K
30 <=50K
31 >50K
32 >50K
33 <=50K
34 <=50K
35 >50K
36 >50K
37 <=50K
38 <=50K
39 >50K
```