

Laporan Pengerjaan Tugas 1
Naive Bayess



Mata Kuliah :

Machine Learning

Irvan Naufali Rahmanto
IF – 40 - 08

A. Analisa masalah

Pada permasalahan yang terdapat ini, kita dapatkan sebuah masalah dimana terdapat sebuah trainset himpunan data berisi 160 objek data yang memiliki 7 atribut dan 1 output yang memiliki 4 kelas/label. Dari sana kita dapatkan sebuah teori penyelesaian tersebut menggunakan metode *naive bayes*, selanjutnya metode *naive bayes* itu sendiri ialah metode dimana memanfaatkan perhitungan probabilitas dan statistik yaitu memprediksi sebuah probabilitas di masa depan dari masa sebelum nya.

B. Strategi Penyelesaian

a) Pembacaan file Trainset.csv pada program

Pada program terdapat sebuah file TrainsetTugas1ML.csv kita import terlebih dahulu kepada program, dimana file tersebut harus terpanggil saat di running kemudian file tersebut (csv nya) nanti akan kita masukkan ke dalam Array, yang nama Array nya Arrtrain dan di inisialisasi oleh suatu variabel i.

b) Menghitung Probabilitas kelas

Masuk pada pembuatan sebuah function atau memiliki hasil keluaran (*return*) pada function bernama Kelasprobabilitas ini akan menghitung berapa banyak jumlah $>50K$ dan $\leq 50K$ pada label ke 8 pada csv yang bernama label income.

c) Mencari Probabilitas Atribut

Pada step sebelum nya kita sudah dapatkan berapa banyak atau jumlah dari label income yang >50 dan $\leq 50K$, selanjut kita harus dapatkan berdasarkan label label lain nya mulai dari age hingga hoursperweek yang berjumlah 7 label berapa banyak yang >50 dan ≤ 50 per label nya.

d) Mencari Probabilitas atribut yang $>50K$

Perhitungan mencari berapa banyak yang >50 dan ≤ 50 per label nya dimana dengan pengecekan kondisi ditiap variabel/atribut masing-masing label nya dibanding kan dengan label income pada index ke 7 atau yg >50 .

e) Mencari Probabilitas atribut yang $\leq 50K$

Perhitungan mencari berapa banyak yang >50 dan ≤ 50 per label nya dimana dengan pengecekan kondisi ditiap variabel/atribut masing-masing label nya dibanding kan dengan label income pada index ke 7 atau yg ≥ 50 .

f) Pemanggilan data Testset.csv pada program

Dari hasil perhitungan masing masing probabilitas bersertakan perhitungan kelas nya (income) mana yang >50 dan ≤ 50 maka selanjut nya kita tinggal

lakukan tes menggunakan data test yang sudah disediakan berupa TestsetTugas1ML.csv dengan memanggil file tsb kemudian kita buat pengecekan kondisi pada function atributProbabilitasBsr50 dengan atributProbabilitasKcl50.

C. Hasil Output

a) Hasil akurasi tebakan

	A		
1	<=50K	21	>50K
2	<=50K	22	>50K
3	>50K	23	>50K
4	<=50K	24	>50K
5	>50K	25	>50K
6	>50K	26	>50K
7	<=50K	27	>50K
8	<=50K	28	>50K
9	>50K	29	>50K
10	>50K	30	<=50K
11	>50K	31	<=50K
12	>50K	32	<=50K
13	<=50K	33	>50K
14	>50K	34	>50K
15	>50K	35	<=50K
16	>50K	36	>50K
17	<=50K	37	<=50K
18	>50K	38	>50K
19	<=50K	39	>50K
20	>50K	40	>50K
21	>50K		