

## Laporan Pengerjaan Tugas 1 Naive Bayess

Mata Kuliah : Machine Learning

Ainun Abidin – 1301164159 – if 40 08

### Persoalan

Diberikan sebuah Trainset berupa himpunan data berisi 160 objek data yang memiliki 7 atribut input (**age, workclass, education, marital-status, occupation, relationship, hours-per-week**) dan 1 output (label kelas **income**) yang memiliki 2 kelas/label ( $>50K$ , dan  $\leq 50K$ ). Bangunlah sebuah sistem klasifikasi menggunakan metode **Naïve Bayes** untuk menentukan kelas/label data testing dalam Testset. Sistem membaca masukan file TrainsetTugas1ML.csv dan TestsetTugas1ML.csv dan mengeluarkan *output* berupa file **TebakanTugas1ML.csv** berupa satu kolom berisi **40 baris** yang menyatakan kelas/label baris yang bersesuaian pada file TestsetTugas1ML.csv.

### Analisis dan strategi

Terdapat data train yang dilakukan perhitungan probabilitas nya dan kemudian akan di perbandingkan dengan data test dan dihitung probabilitas nya dan diberi label dengan nilai yang terbaik

### Penyelesaian

Pertama dilakukan adalah melakukan persiapan data dengan cara memanggil dua file csv yang bernama

TrainsetTugas1ML.csv sebagai data train dan TestsetTugas1ML.csv sebagai data testnya dan memasukan kedua data tersebut kedalam dua variable train dan test

Kedua adalah melakukan perhitungan probabilitas kelas atau labelnya dalam peraoalan kali ini terdapat 2 label Yaitu  $>50K$  dan  $\leq 50k$  perhitungan dilakukan dan  $>50k$  saya perumakan menjadi variable terima dan  $\leq 50k$  sebagai variable tidak dan dihitung probabilitasnya

Probabilitas terima =  $\frac{\text{nterima}}{(\text{nterima} + \text{ntidak})}$

Probabilitas tidak =  $\frac{\text{ntidak}}{(\text{nterima} + \text{ntidak})}$

Ketiga adalah melakukan perhitungan probabilitas  $>50K$  dan  $\leq 50k$  terhadap data test dengan cara memperbandingkan class dari train yang mempunyai variable yang sama dan memiliki label yang sama

Probabilitas class terima terhadap data test( $>50K$ ) =  $(\text{nclas yang sama} / \text{terima} * \text{nclass yang lain yang sama} / \text{terima} \dots * \text{probabilitas terima})$

Dan melakukan perhitungan class tidak terhadap data test ( $\leq 50K$ ) =  $(\text{nclas yang sama} / \text{tidak} * \text{nclass yang lain yang sama} / \text{tidak} \dots * \text{probabilitas tidak})$

Keempat melakukan perbandingan untuk menentukan dan memberikan label

Dengan cara

Jika Probabilitas class terima terhadap data test  $>$  Probabilitas class tidak terhadap data test

Maka menghasilkan label = ( $>50K$ )

Jika Jika Probabilitas class terima terhadap data test  $<$  Probabilitas class tidak terhadap data test

Maka menghasilkan label = ( $\leq 50K$ )

Maka menghasilkan tebakan dari tugas dalam klasifikasi naïve bayes

Hasil Tebakan

	A	B
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	
40	>50K	
41	Accuracy: 97.5609756097561	
42		

Untuk menghitung akurasi adalah jumlah data yang berhasil ditebak

$$\frac{\text{Ntebakan}}{n(\text{length test})} \times 100$$