

Nama : Fuad Zauqi Nur

Kelas : IF-40-03

NIM : 1301164392

---

## **I. Deskripsi Masalah**

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.

## **II. Strategi Penyelesaian**

Algoritma Terlampir pada tugas1ml.py

Penjelasan Algoritma :

1. Mengimport modul pandas agar mudah membaca csv file.
2. Train\_r adalah jumlah dari himpunan data train yang didapat dari TrainsetTugas1ML.csv
3. Test\_r adalah jumlah dari himpunan data train yang didapat dari TestsetTugas1ML.csv..
4. Plus adalah jumlah dari data yang memiliki nilai peluang lebih tinggi.
5. Minus adalah jumlah dari data yang memiliki nilai peluang rendah.
6. Naivebayes(col1,cat1,col2,cat2,tipe) adalah persamaan untuk menghitung peluang dari semua data yang ada di data test.
7. Plus\_p adalah perbandingan antara jumlahdata yang memiliki nilai peluang tinggi dan jumlah data train keseluruhan.
8. Plus\_p adalah perbandingan antara jumlahdata yang memiliki nilai peluang rendah dan jumlah data train keseluruhan
9. P\_plus\_... adalah nilai peluang dari data peluang bernilai tinggi dan P\_minus... adalah sebaliknya.
10. Pp dan pm adalah persamaan penentuan priority.

Penjelasan penggunaan program :

1. Install python versi 3.7.0 agar memiliki versi python yang sama dengan program yang dibuat
2. Install pandas agar dapat memanggil atau mengimport modul pandas
3. Install xlrd agar dapat membaca file excel atau csv
4. Eksekusi tugas1ml.py maka akan menghasilkan data yang disimpan pada TebakanTugas1ML.csv

Hasil Running :

[illegible]

Hasil TebakanTugas1ML.csv :

	A	B
1	id	income
2	26027	>50K
3	26314	>50K
4	31405	<=50K
5	14736	>50K
6	27217	<=50K
7	5951	<=50K
8	30067	>50K
9	28177	>50K
10	15390	<=50K
11	18042	<=50K
12	5793	<=50K
13	31274	<=50K
14	17068	>50K
15	21834	<=50K
16	24128	<=50K
17	8550	<=50K
18	1181	>50K
19	11149	<=50K
20	20836	>50K
21	25766	<=50K
22	139	<=50K
23	27160	<=50K
24	8814	<=50K
25	11470	<=50K
26	16634	<=50K
27	26306	<=50K
28	9830	<=50K
29	3763	<=50K
30	11078	<=50K
31	12324	>50K
32	26020	>50K
33	3017	>50K
34	10243	<=50K
35	10882	<=50K
36	19355	>50K
37	4565	<=50K
38	23102	>50K
39	26673	>50K
40	23415	<=50K
41	17337	<=50K