

Analisis Masalah

Terdapat 160 data train yang berisi 7 atribut input dan 1 output yang memiliki 2 label yaitu ">50K" dan "<=50K", lalu terdapat 40 data test yang berisi 7 atribut input dan 1 output yang belum ditentukan labelnya. Akan ditentukan label dari setiap data test dengan membaca dari data train menggunakan metode *naive bayes*. 40 label dari data test akan dioutputkan berupa file TebakanTugas1ML.csv.

Strategi Penyelesaian Masalah

Untuk menentukan label menggunakan metode *Naive Bayes* terdapat beberapa step yaitu :

Step 1 : menghitung probabilitas $P(\text{income} = ">50K")$ dan $P(\text{income} = "<=50K")$

Nilai $P(\text{income} = ">50K")$ dan $P(\text{income} = "<=50K")$ adalah nilai banyaknya kelas *income* dibagi dengan banyaknya data pada data train.

```
static double plebih50(String[][] train){
    int a=0;
    for (int i = 0; i < train.length; i++) {
        if (">50K".equals(train[i][8])) a++;
    }
    return (double) a/train.length;
}
```

```
double plebih = plebih50(t);
double pkurang = 1-plebih50(t);
```

Step 2 : menghitung probabilitas $P(x | ">50K")$ dan $P(x | "<=50K")$

Nilai $P(x | ">50K")$ dan $P(x | "<=50K")$ adalah probabilitas dari setiap atribut pada data yang diuji. Lalu mengalikan setiap probabilitas dari atribut yang memiliki kelas *income* yang sama.

```
static double patribut(String a,String c, int b,double p, String[][] train){
    int z=0;
    for (int i = 0; i < train.length; i++) {
        if ((train[i][b].equals(a)) && (train[i][8].equals(c))) z++;
    }
    return (double) z/(p*train.length);
}

for (int j = 1; j < tes.length; j++) {
    q=q*patribut(tes[j], ">50K", j,plebih,t);
    r=r*patribut(tes[j], "<=50K", j,pkurang,t);
}
```

Step 3 : Mengalikan $P(\text{income}=y)$ dengan $P(\text{atribut}=x \mid \text{income}=y)$

Step ini untuk mendapatkan nilai $P(\text{income}=">50K" \mid x)$ dan $P(\text{income}="<=50K" \mid x)$

```
q=q*plebih;  
r=r*pkurang;
```

Step 4 : membandingkan nilai $P(\text{income}=">50K" \mid x)$ dan $P(\text{income}="<=50K" \mid x)$

Setelah mendapatkan nilai $P(\text{income}=">50K" \mid x)$ dan $P(\text{income}="<=50K" \mid x)$, lalu dibandingkan jika $P(\text{income}=">50K" \mid x)$ lebih besar maka objek data memiliki label ">50K" dan sebaliknya. Label kelas dimasukkan ke dalam file yang bernama TebakanTugas1ML.csv.

```
static String choosemax(double a, double b){  
    if (a>b) return ">50K";  
    else return "<=50K";  
}  
  
try{  
    fw.append(String.valueOf(choosemax(q,r))+'\n');  
}catch(Exception e){}  
}  
fw.flush();  
fw.close();
```

Hasil Running :

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