



**LAPORAN TUGAS**

# **Sistem Klasifikasi Menggunakan Metode Naïve Bayes untuk Menentukan Kelas Data Testing**

Mata Kuliah

Machine Learning

Disusun Oleh:

Muhammad Aditya Rayhan 1301160438

Telkom University

Bandung

2019

## 1. Abstraksi

Naïve Bayes adalah sebuah probabilistic classifier dalam machine learning yang menerapkan teorema bayes dengan asumsi independent naive di setiap fitur pada data. Dalam kasus ini, Algoritma Naïve Bayes digunakan untuk menyelesaikan penentuan kelas data test melalui data training yang sudah diberikan.

## 2. Pendahuluan

Naïve Bayes adalah sebuah probabilistic classifier dalam machine learning yang menerapkan teorema bayes dengan asumsi independent naive di setiap fitur pada data. Naïve Bayes memiliki teorem umum seperti berikut,

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

$P(A|B)$  merupakan probabilitas kondisional pada terjadinya kejadian A ketika kejadian B bersifat true.  $P(B|A)$  merupakan probabilitas kondisional pada terjadinya kejadian B ketika kejadian A bersifat true.  $P(A)$  merupakan probabilitas banyaknya kejadian A yang muncul pada data.  $P(B)$  merupakan probabilitas banyaknya kejadian B yang muncul pada data. Namun  $P(B)$  tidak dipakai karena pada kasus ini hanya dibutuhkan perkalian  $P(B|A)$  dan  $P(A)$  saja sehingga didapatkan nilai untuk perbandingan.

Kasus yang dihadapi pada pembahasan ini adalah 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.

## 3. Penjelasan Algoritma

### 3.1 Input dan Output Algoritma

Algoritma Naïve Bayes diawali membuka data csv sehingga didapat Data Train dan Data Test. Data Train diperoleh dari TrainsetTugas1ML.csv dan Data Test diperoleh dari TestsetTugas1ML.csv. Pada output, algoritma ini akan mengeluarkan file TebakanTugas1ML.csv yang berisi kelas dari Data Test yang sudah diklasifikasi dengan model yang dibangun dari Data Train.

### 3.2 Cara Kerja Algoritma

#### Analisis dan Strategi Penyelesaian Masalah

Data train yang sudah didapat, masing masing fiturnya dihitung dan dijumlahkan. dibagi kedalam 2 variable berbeda berdasarkan kepemilikan labelnya. Variabel countup ditujukan pada data yang memiliki label ' $>50K$ ' dan countdown ditujukan pada data yang memiliki label ' $\leq 50K$ '.

Setelah itu dilakukan perhitungan pada data test menggunakan teorema bayes dengan melibatkan fitur yang ada pada data test seperti fitur age, workclass, education, marital-status, occupation, relationship, dan hours-per-week lalu kemudian dibandingkan hasilnya pada kondisi kejadian dengan label '>50K' dan '<=50K' yang sudah didapatkan dari data train. Hasil yang lebih baik pada suatu label maka labelnya akan menjadi label pada data test.

Setelah semua data test mendapat label, label data test disimpan pada `TebakanTugas1ML.csv` yang merupakan output dari algoritma Naive Bayes.

```
In [3]: runfile('E:/Machine Learning/Naive Bayes/1301160438.py', wdir='E:/Machine Learning/Naive Bayes')
```

A
<=50K
<=50K
>50K
<=50K
>50K
>50K
<=50K
<=50K
>50K
>50K
>50K
>50K
>50K
<=50K
>50K