

LAPORAN TUGAS

Sistem Klasifikasi Menggunakan Metode Naïve Bayes untuk Menentukan Kelas Data Testing

Mata Kuliah

Machine Learning



Disusun Oleh :

M. Naufal Syawali . A

1301164488

S1 Teknik Informatika

Telkom University

Bandung

2019

1. Abstraksi

Naïve Bayes adalah salah satu metode Pembelajaran Mesin (Machine Learning) yang memanfaatkan perhitungan probabilitas dan statistika, yaitu dengan memprediksi probabilitas atau peluang di masa depan menggunakan dasar pengalaman di masa sebelumnya, dalam kasus ini Algoritma Naïve Bayes digunakan untuk menyelesaikan kelas data set melalui data training yang sudah ada/diberikan.

2. Pendahuluan

Naïve Bayes adalah sebuah probabilitas classifier dalam machine learning yang menerapkan teorema bayes dengan asumsi independent naïve di setiap fitur pada data. Naïve Bayes memiliki teorem/persamaan umum seperti berikut :

Bayes theorem

$$P(h | D) = \frac{P(D | h)P(h)}{P(D)}$$

Penjelasan dari Naïve bayes teorem diatas :

$P(h|D)$ = kondisi probabilitas dari h terhadap D.

$P(D|h)$ = kondisi probabilitas dari D terhadap h (kemungkinan D terhadap h atau likelihood).

$P(D)$ = independen probabilitas dari D.

$P(h)$ = independen probabilitas dari h (prioritas kemungkinan).

Kasus yang dihadapi pada pembahasan ini adalah diberikan sebuah Trainset berupa data sebanyak 160 data objek dalam sebuah table yang memiliki 7 kolom yaitu age, workclass, education, marital-status, occupation, relationship, hours-per-week. Dan 1 output di dalam kolom ke 8 atau kolom terakhir yaitu Income yang memiliki 2 kemungkinan label/status yang diberikan (>50K atau <=50K).

Perintahnya adalah membangun sebuah sistem klasifikasi menggunakan metode Naïve Bayes untuk menentukan kelas/label data testing. Sistem membaca masukkan data train dari file yang sudah di berikan yaitu TrainerTugas1ML.csv dan TestsetTugas1ML.csv dan mengeluarkan atau memberikan hasil tebakan dari data hasil Test final, berupa satu kolom berisi 40 baris yang menyatakan kelas/label baris yang bersesuaian pada file TestsetTugas1ML.csv

3. Penjelasan Algoritma

Input dan Output Algoritma

Langkah pertama yang dilakukan dalam pembuatan sistem ini adalah dengan mengimport data TrainsetTugas1ML.csv dan TestsetTugas1ML.csv dengan menggunakan algoritma Naïve bayes yang diawali dengan membuka 2 file tersebut,

dan pada hasil akhir program atau sistem mengeluarkan output berupa DataTebakanTugas1.csv seperti berikut

	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		