

Pembuatan Sistem Klasifikasi Menggunakan Naïve Bayes Untuk Menentukan Label Kelas Income

Muh. Hasbi Abdul M – 1301160335

Program Studi S1 Informatika, Fakultas Informatika
Universitas Telkom Bandung
Jl. Telekomunikasi Dayeuhkolot, Bandung 40257

III. ANALISIS PERMASALAHAN

I. PENDAHULUAN

Naïve Bayes merupakan sebuah metode klasifikasi yang menggunakan probabilitas dan statistik. Algoritma ini memprediksi kelas data baru (dataTest) berdasarkan perhitungan pada data lama (dataTrain). Ciri utama dari algoritma ini adalah independensi asumsi yang sangat kuat (Naïve) dari setiap kondisi yang ada. Kelebihan dari algoritma ini adalah metode ini tidak memerlukan jumlah dataTrain yang banyak untuk menentukan estimasi parameter yang dibutuhkan saat proses pengklasifikasian, selain itu algoritma ini juga mudah dibuat dan hasilnya juga bagus. Untuk kekurangannya sendiri yaitu asumsi independence antar atribut membuat akurasi berkurang.

II. MASALAH

Menentukan kategori / kelas (income) dari sebuah file TestsetTugas1ML.csv yang berisikan 40 data dengan beberapa atribut yaitu age, workclass, education, marital-status, occupation, relationship, hours-per-week. Dengan algoritma Naïve Bayes diminta untuk membuat sistem yang dapat menentukan kelas(income) dari 40 data tersebut berdasarkan data pada TrainSetTugas1ML.csv yang berisikan 160 data.

Pada metode ini terdapat beberapa tahap yang harus dilakukan, yaitu

Sebelum memulai tahap pertama hal yang harus dilakukan adalah memasukkan *data train & data test* yang ada ke dalam program.

Tahap pertama yaitu menghitung jumlah kelas, pada kasus ini terdapat 2 kelas yaitu data dengan income besar (>50K) dan data dengan income kecil (<=50K). Setelah itu saya memisahkan dataTrain berdasarkan klasifikasi tersebut menjadi 2 list yaitu list incomeBesar dan list incomeKecil.

Tahap kedua menghitung frekuensi / jumlah kasus per kelas / atribut. Karena pada data ini terdapat 8 atribut dan setiap atribut memiliki 3 klasifikasi maka yang saya lakukan adalah menghitung satu per-satu setiap klasifikasi dalam suatu atribut. Misalkan pada kasus ini saya menghitung frekuensi atribut age terlebih dahulu maka saya menghitung jumlah setiap klasifikasi yang ada pada atribut age baik yang income nya >50K maupun yang <=50K dengan menamainya sesuai klasifikasi atribut age (ex : adultBesar untuk jumlah kelas age dengan klasifikasi adult yang memiliki income besar).

Tahap ketiga menghitung jumlah data dengan income kecil & besar yang ditampung pada variabel jumBesar dan jumKecil dan menghitung probabilitas income besar dan income kecil dengan membagi jumlah data income kecil / besar dengan banyak data pada dataTrain.

Tahap keempat menghitung probabilitas tiap atribut yang ditampung dalam sebuah dictionary untuk setiap atribut yang berisikan probabilitasnya dan karena probabilitas setiap kelas tidak ada yang bernilai 0 maka tidak diperlukan *smoothing*. Contoh dibawah ini merupakan dictionary untuk menampung probabilitas dari atribut age

```
{'adult': {'<=50k': 0.475, '>50k': 0.44166666666666665},  
'old': {'<=50k': 0.025, '>50k': 0.008333333333333333},  
'young': {'<=50k': 0.5, '>50k': 0.55}}
```

Gambar 1 : Contoh Dictionary Atribut Age

Tahap kelima adalah menentukan kelas dengan mengalikan probabilitas semua atribut pada dataTest dengan probabilitas kelas (income) yang pada program ditampung menggunakan variabel probtotalKecil untuk probabilitas total dengan income kecil pada 1 data dan probtotalBesar untuk probabilitas total dengan income Besar. Setelah mendapat nilai probtotalKecil & probtotalBesar maka langkah selanjutnya yaitu membandingkan kedua nilai tersebut, jika nilai probtotalBesar lebih besar dibanding probtotalKecil maka dapat ditentukan bahwa kelas(income) dari data tersebut adalah >50K dan begitu juga sebaliknya.

Tahap terakhir yaitu menuliskan hasil klasifikasi kedalam file baru bernama TebakanTugas1ML.csv yang isinya hanya berupa 1 kolom berisi 40 baris yang menyatakan klasifikasi yang telah dilakukan.

IV. REFERENSI

[1] Informatikalogi. 8 April 2017. "Algoritma Naive Bayes". <https://informatikalogi.com/algoritma-naive-bayes/>

[2] Slide Kuliah Pembelajaran Mesin Fakultas Informatika Telkom University "Naive Bayes"