Laporan Pengerjaan Tugas 1 Naive Bayess

Mata Kuliah: Machine Learning

Ainun Abidin – 1301164159 – if 40 08

Persoalan

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 <=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.

Analisis dan strategi

Terdapat data train yang dilakukan perhitungan probabilitas nya dan kemudian akan di perbandingkan dengan data test dan dihitung probabilitas nya dan diberi label dengan nilai yang terbaik

Penyelesaian

Pertama dilakukan adalah melakukan persiapan data dengan cara memanggil dua file csv yang bernama

TrainsetTugas1ML.csv sebagai data train dan TestsetTugas1ML.csv sebagai data testnya dan memasukan kedua data tersebut kedalam dua variable train dan test

Kedua adalah melakukan perhitungan probabilitas kelas atau labelnya dalam peroalan kali ini terdapat 2 label Yaitu >50K dan <= 50 k perhitungan dilakukan dan >50k saya perumakan menjadi variable terima dan <=50k sebagai variable tidak dan dihitung probabilitasnya

Probabilitas terima = nterima / (nterima+ntidak)

Probabilitas tidak = ntidak / (nterima+ntidak)

Ketiga adalah melakukan perhitungan probabilitas >50K dan <=50k terhadap data test dengan cara memperbadingkan class dari train yang mempunyai variable yang sama dan memiliki label yang sama

Probabilitas class terima terhadap data test(>50K) = (nclas yang sama / terima *nclass yang lain yang sama / terima.... *probabilitas terima

Dan melakukan perhitungan class tidak terhadap data test (<=50K) = = (nclas yang sama /tidak *nclass yang lain yang sama / tidak *probabilitas tidak

Keempat melakukan perbandingan untuk menetukan dan memberikan label

Dengan cara

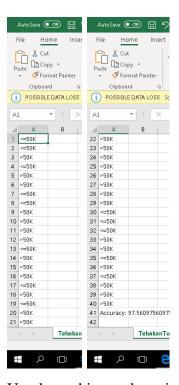
Jika Probabilitas class terima terhadap data test > Probabilitas class tidak terhadap data test

Maka mengahasilkan label = (>50K)

Jika Jika Probabilitas class terima terhadap data test < Probabilitas class tidak terhadap data test Maka menghasilkan label = (<=50K)

Maka mengahsilkan tebakan dari tugas dalam clasifikasi naïve bayes

Hasil Tebakan



Untuk menhitung akurasi adalah jumlah data yang berhasil ditebak

Ntebakan/n(length test) x 100