**Tugas 4, *Machine Learning***

**Laporan Membangun Sistem Klasifikasi Untuk Menentukan Kelas/Label Data Uji Menggunakan Metode *Bagging* dari *Ensemble Learning Berbasis* *Naïve Bayes***

Oleh:

Enrico Farizky Rustam ( 1301164263 )

IF 40-04 / S1 Informatika / Universitas Telkom

**Abstrak**: Himpunan data berisi 298 objek data yang memiliki 2 atribut *input* (**X1** dan **X2**) dan 1 *output* (**Class**) yang memiliki dua label (bernilai 1 atau 2) dengan menggunakan Metode *Bagging* yaitu salah satu teknik dari *Ensamble Learning* berbasis *Naïve Bayes* . Metode *Naïve Bayes Classifiers* yaitu salah satu metode klasifikasi teks berdasarkan probabilitas kata kunci dalam mengambil beberapa sample dari data uji yang disediakan, kemudian diberi label dan dicek apakah data sample tersebut sesuai dengan data uji atau tidak. Selanjutnya, sistem akan membaca masukan file data latih dan data uji dan mengeluarkan *output* berupa file satu kolom berisi **75 baris** yang menyatakan kelas/label baris yang bersesuaian pada file data uji.

**Kata Kunci** : *Bagging, Ensamble Learning*, *Naïve Bayes*, data latih, data uji, probabilitas

# Pendahuluan

Pengklasifikasi bayes merupakan salah satu pengklasifikasi statistik, dimana pengklasifikasi ini dapat memprediksi probabilitas keanggotaan kelas suatu data yang akan masuk ke dalam kelas tertentu, sesuai dengan perhitungan probabilitas. Pengklasifikasi Bayes didasari oleh teorema bayes yang ditemukan oleh Thomas Bayes pada abad ke-18. *Naïve Bayes Classifier* menunjukkan akurasi dan kecepatan yang tinggi bila diterapkan pada database yang besar. Metode ini sering digunakan dalam menyelesaikan masalah dalam bidang mesin pembelajaran karena metode ini dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana.

Teorema bayes merupakan formula dasar aturan dari *Naïve Bayes Classifier* dengan formula umum sebagai berikut:

A drawing of a person

Description generated with high confidence

*Bagging (Bootstrap aggregating)* ditemukan oleh Leo Breiman pada 1994. Merupakan metode

*ensemble learning* efektif yang pertama kali diusulkan. Dikenal sebagai salah satu metode yang paling simpel dalam *arching* (*adaptive reweighting and combining*), yang mengacu pada pemilihan data untuk meningkatkan akurasi klasifikasi (Sewell 2018).

Awal kemunculannya, *bagging* dirancang untuk klasifikasi berbasis *decision tree.* Tetapi, dalam perkembangannya *bagging* dapat digunakan untuk berbagai model, seperti *Naïve Bayes, ANN, SVM,* dan sebagainya, dan dapat digunakan untuk klasifikasi maupun regresi.

# Deskripsi Soal Masalah

Masalah dari soal ini yaitu membangun sebuah kondisi, dimana file yang diberikan yaitu file TrainsetTugas4ML.csv berupa himpunan data berisi 298 objek data yang memiliki 2 atribut input (**X1** dan **X2**) 1 output (**Class**) yang memiliki 2 label (bernilai 1 atau 2) menggunakan Metode *Bagging* yaitu salah satu teknik dari *Ensamble Learning* berbasis *Naïve Bayes* . Kemudian, dengan menggunakan metode *Naïve Bayes*, harus menentukan kelas/label data testing dalam file yang diberikan.

Sistem memasukkan file masukan file TrainsetTugas4ML.csv dan membaca data dalam file tersbut dan mengeluarkan *output* be rupa file **TebakanTugas4ML.csv** berupa satu kolom berisi **75 baris** yang menyatakan kelas/label baris yang bersesuaian.

# Metode Penyelesaian

Algoritma *Naïve Bayes**Classifier* (NBC) merupakan salah satu metoda [pembelajaran mesin](https://id.wikipedia.org/wiki/Pemelajaran_mesin) yang memanfaatkan perhitungan proba-bilitas dan statistik yang memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Secara umum pada soal ini, mendefinisikan menghitung data acak dari data sample yg diambil kemudian diklasifikasikan menggu-nakan metode Naive Bayes dan diambil jumlah data yang cocok lalu diberikan label sesuai dengan data uji dan latih yang ada. Setelah hasilnya didapat, kemudian dijumlahkan sebagai berikut:

def count(x, rata2, stdev):

var = float(stdev)\*\*2

pembagi = (2\*math.pi\*var)\*\*.5

num = math.exp(-(float(x)-float(rata2))\*\*2/(2\*var))

return num/pembagi

def nvBayes(DU, mnstdv): #naive bayes

kls1 = count(DU[0], mnstdv[0][0], mnstdv[0][1]) \* count(DU[1], mnstdv[2][0], mnstdv[2][1]) \* (mnstdv[0][2]/dttrainmdl)

kls2 = count(DU[0], mnstdv[1][0], mnstdv[1][1]) \* count(DU[1], mnstdv[3][0], mnstdv[3][1]) \* (mnstdv[1][2]/dttrainmdl)

if(kls1>kls2):

return -1

else:

return 1

# Parameter & Fungsi Klasifikasi

Pada program *Naïve Bayes**Classifier* ini saya menggunakan 2 data inputan, yaitu **X1** dan **X2.** Lalu memiliki 1 output (**Class)** yang memiliki 2 label (bernilai 1 atau 2). Data Tes yang nilai kelas sudah dan data Train data kelas belum ditentukan. Fungsi statistics untuk menghitung standar deviasi.

def meanstdev(bt): #fungsi hitung rata2 dan standar deviasi

attr1 = []; attr2 = []; attr3 = []; attr4 = []

jml1 = 0; jml2 = 0

panjang=len(bt)

for i in bt:

if(i[2]==1):

attr1.append(i[0])

attr3.append(i[1])

jml1+=1

elif(i[2]==2):

attr2.append(i[0])

attr4.append(i[1])

jml2+=1

hsl1 = [statistics.mean(attr1), statistics.stdev(attr1), jml1]

hsl2 = [statistics.mean(attr2), statistics.stdev(attr2), jml2]

hsl3 = [statistics.mean(attr3), statistics.stdev(attr3), jml1]

hsl4 = [statistics.mean(attr4), statistics.stdev(attr4), jml2]

mnstdev = []

mnstdev.append(hsl1); mnstdev.append(hsl2); mnstdev.append(hsl3); mnstdev.append(hsl4)

return mnstdev

Fungsi Klasifikasi digunakan untuk menentukan kelas/label dari data uji yang dicari dengan algoritma perulangan Naive Bayes, lalu diurutkan berdasarkan peluang yang benar sesuai dengan data uji yang ada.

# Output Program

Output program disimpan dalam bentuk csv dengan nama file TebakanTugas4ML.csv dan dengan indeks sebagai urutan satu kolom berisi 40 baris yang menyatakan kelas income baris yang bersesuaian.



# Sumber

# <https://id.wikipedia.org/wiki/Naive_Bayes_classifier>

1. Septian,Yuda. “DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI KELULUSAN MAHASISWA UNIVERSITAS DIAN NUSWANTORO”. Jurusan Sistem Informasi, Fakultas Ilmu Komputer.