

Problem :

Diberikan sebuah Trainset berupa 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). Bangunlah sebuah sistem klasifikasi menggunakan metode *Bagging*, salah satu teknik *Ensemble Learning*, berbasis *Naïve Bayes* untuk menentukan kelas/label data uji dalam Testset. Sistem membaca masukan file TrainsetTugas4ML.csv dan TestsetTugas4ML.csv dan mengeluarkan *output* berupa file TebakanTugas4ML.csv berupa satu kolom berisi 75 baris yang menyatakan kelas/label baris yang bersesuaian pada file TestsetTugas4ML.csv.

Strategi Penyelesaian Masalah :

Strategi yang dilakukan yaitu dengan mengimplementasikan metode **Bagging** pada **Ensemble Learning** kedalam sebuah program sehingga program tersebut dapat memberikan hasil kelas kepada masing-masing data yang terdapat di data set. Untuk melakukan **Bagging** saya terlebih dahulu menentukan banyaknya bootstrap yang saya pakai yaitu sebanyak 7 model bootstrap yang berbeda dimana data yang berada di dalam setiap model bootstrap saya dapatkan dari random data pada data train dan banyak data yang terdapat pada masing-masing bootstrap sebanyak 298 data. Setelah sudah mendapatkan model bootstrap maka selanjutnya saya mempredict setiap data di data set dengan masing-masing model bootstrap dan di simpan ke dalam 7 buah array untuk keluaran semua data di 7 buah model bootstrap, untuk predict data saya menggunakan library naive bayes. Setelah sudah mendapatkan hasil dari predict data terhadap masing-masing model bootstrap, selanjutnya setiap element akan program cek data pada element tersebut kebanyakan hasil 1 atau hasil 2 , jika lebih banyak hasil 1 maka data pada element tersebut memiliki class/label 1 , jika lebih banyak hasil 2 maka data pada data set di element tersebut akan memiliki class/label 2.

Analisis:

Berdasarkan penjelasan yang saya jabarkan maka dapat diketahui bahwa banyaknya pemakaian model bootstrap sangat berperan penting dalam mendapatkan akurasi yang lebih besar dan pemilihan data random yang berada di dalam model bootstrap juga berperan penting untuk mendapatkan predict yang sesuai.

Hasil Percobaan :

```

total += 1
print('Akurasi')
print(str((total/298)*100) + ' %')

```

Akurasi
94.6308724832 %

1	Class	32	2	70	1
2	2	33	1	71	1
3	2	34	2	72	1
4	2	35	1	73	1
5	2	36	1	74	1
6	2	37	1	75	1
7	2	38	1	76	1
8	2	39	1		
9	2	40	1		
10	2	41	1		
11	2	42	1		
12	2	43	2		
13	1	44	1		
14	2	45	1		
15	1	46	1		
16	1	47	1		
17	1	48	1		
18	1	49	1		
19	1	50	1		
20	1	51	1		
21	1	52	1		
22	2	53	1		
23	1	54	1		
24	2	55	1		
25	1	56	1		
26	1	57	1		
27	1	58	1		
28	1	59	1		
29	2	60	1		
30	1	61	1		
31	2	62	1		
		63	1		
		64	1		
		65	1		
		66	1		
		67	1		
		68	1		
		69	1		