

Nama : Esy Santoso

Nim : 1301162111

Kelas : IF-4008

Laporan Klasifikasi Data dengan menggunakan Metode bagging

Deskripsi Masalah

Terdapat 2 dataset yang diberikan berupa data train data test. Data train yang digunakan mempunyai 298 data sampel yang memiliki 2 atribut dan 1 kelas. Sedangkan data test terdapat 2 atribut dan 1 kelas yang belum terdefinisi. Diperlukan sebuah sistem untuk memprediksi kelas pada dataset. Dalam hal ini akan digunakan sebuah metode bagging dengan naïve bayes sebagai metode klasifikasinya.

Metode Penyelesaian

Permasalahan yang telah disebutkan diatas, akan diselesaikan dengan menggunakan bagging. Bagging adalah metode yang dapat memperbaiki hasil dari algoritma klasifikasi mesin learning. Dengan adanya Bagging, ketidakstabilan model dapat diatasi.

Akan dibuat 17 model dengan 200 data disetiap modelnya. Data yang akan dipakai berasal dari data train dan dipilih secara random. Cara kerja dari bagging dan digabungkan dengan naïve Bayes dengan classifier sebagai berikut:

1. buat sebanyak 17 model, dimana setiap modelnya terdiri dari 200 data yang pilih secara random.
2. setelah itu, lakukan perhitungan klasifikasi naïve bayes, yang mana cara kerja nya seperti berikut:
 - a. Pisahkan data yang mempunyai kelas 1 dan 2
 - b. setelah itu terapkan rumus berikut ini

$$P(x_k | C_i) = \frac{1}{\sigma_{ik} \sqrt{2\pi}} e^{-\frac{(x_k - \mu_{ik})^2}{2\sigma_{ik}^2}}$$

Dimana σ adalah standar deviasi dari data.

- c. setelah itu, bandingkan antara perhitungan untuk kelas 1 dan 2, ambil kelas yang paling besar
3. lakukan perhitungan naïve bayes kesemua model yang telah dibuat.
 4. setelah itu, akan dihasilkan gambaran tabel sebagai berikut:

Model 1	X1	6.5	12.2	6.4
	X2	1.3	12.2	4.5
	Y	1	2	2
Model 2	X1	12	6.3	5.32
	X2	1.6	1.7	9.54

	Y	1	2	1
....				

*data hanya ilustrasi

5. langkah terakhir, hasil dari setiap model akan dibandingkan, mana yang terbanyak itulah kelas yang terpilih.

Hasil Pemrosesan

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

Hasil dari pemrosesan klasifikasi dapat dilihat dari gambar diatas. Pemilihan banyaknya model dan banyaknya data yang digunakan akan sangat berpengaruh dengan hasil yang dihasilkan.