

Klasifikasi Data dengan Menggunakan Metode Bagging

Badrus Shoolehk Al Ar Fanny

IF-40-04

Fakultas Informatika, Universitas Telkom

Jl. Telekomunikasi no.1 Terusan Buah Batu, Kab. Bandung, Jawa Barat

badrussholehaxel@gmail.com

Deskripsi Masalah

Diberikan sebuah datatrain yang mempunyai 2 atribut input (**X1** dan **X2**) dan 1 output (**Class**) yang memiliki dua label (bernilai 1 atau 2). Selain itu diberikan juga data test yang terdiri dari 2 atribut **X1** dan **X2** yang mana tidak mempunyai label class. Oleh karena itu diperlukan sebuah sistem untuk memprediksi class yang terdapat pada datatest yang telah diberikan dengan menggunakan metode Bagging.

Metode Penyelesaian

Permasalahan yang telah dituliskan diatas, akan diselesaikan dengan sebuah sistem dengan metode bagging. Bagging merupakan metode yang dapat memperbaiki hasil dari algoritma klasifikasi machine learning untuk mengatasi ketidakstabilan pada model yang dibuat. Pada kasus ini, akan dibuat 15 model data train, yang mana setiap modelnya akan terdiri dari 150 data yang bersala dari data train dan dipilih secara random. Sistem klasifikasi yang akan digunakan dalam setiap modelnya adalah dengan Naïve Bayes.

Adapun langkah kerja yang dilakukan untuk membangun sistem ini sebagai berikut:

1. Tentukan berapa jumlah model yang akan di train dengan menggunakan Naïve Bayes, yang mana dalam kasus ini penulis menggunakan 15 model.
2. Setelah itu, lakukan pemilihan secara random banyaknya data train yang akan digunakan disetiap modelnya. Disetiap model, masing masing akan mempunyai 150 data

yang mana akan dipilih secara random oleh sistem.

3. setelah model telah terbentuk, train model dengan menggunakan metode naïve bayes untuk mengklasifikasi data test yang telah disediakan.
4. proses pengklasifikasian naïve bayes dilakukan dengan menggunakan library sklearn.
5. Lakukan proses klasifikasi kesemua model yang telah dibuat sebelumnya. Jadi proses pengklasifikasian dilakukan sebanyak 15 kali.
6. hasil dari proses klasifikasi kurang lebih akan seperti tabel dibawah ini.

Model 1	X1	12	15.2	2.2
	X2	12.2	13.2	7.8
	Y	1	1	2
Model 2	X1	12	15.2	2.2
	X2	12.2	13.2	7.8
	Y	1	2	2
....				
Model 15				

7. Hasil dari proses klasifikasi diatas, selanjutnya adalah dengan membandingkan kelas mana yang jumlah nya paling banyak untuk setiap di datatest. Hasil kelas yang paling terbanyak, akan menjadi kelas sesungguhnya pada datatest.

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 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	2	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

Gambar disamping adalah hasil pemrosesan dari proses bagging. Dari hasil yang telah didapatkan, pemilihan jumlah model dan jenis data train yang digunakan. Apabila jumlah model yang digunakan diubah, maka hasilnya akan ikut berubah pula.