**LAPORAN ANALISA SENTIMEN LIRIK LAGU**



MATA KULIAH:

INFORMATION RETRIEVAL

Disusun Oleh:

Triariyani (16.01.55.0017)

**Universitas Stikubank Semarang**

**2019**

**Buat laporan analisa sentimen lirik lagu dengan format sebagai berikut :   
Teori :**

1. **Klasifikasi teks bertujuan.....** untuk mengelompokkan teks atau dokumen ke dalam kelas-kelas yang telah didefinisikan sebelumnya
2. **Analisa sentiment adalah....** proses memahami, mengestrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif
3. **Algoritma Naïve Bayes adalah ......** Sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya
4. **Menjelaskan berapa data training dan berapa data test...**

Training-Set ini nantinya akan digunakan untuk membuat model machine learning, sedangkan Test-Set ini akan digunakan untuk menguji performa dan kebenaran (terhadap korelasi) dalam model yang bersangkutan.

Menggunakan Bahasa Python

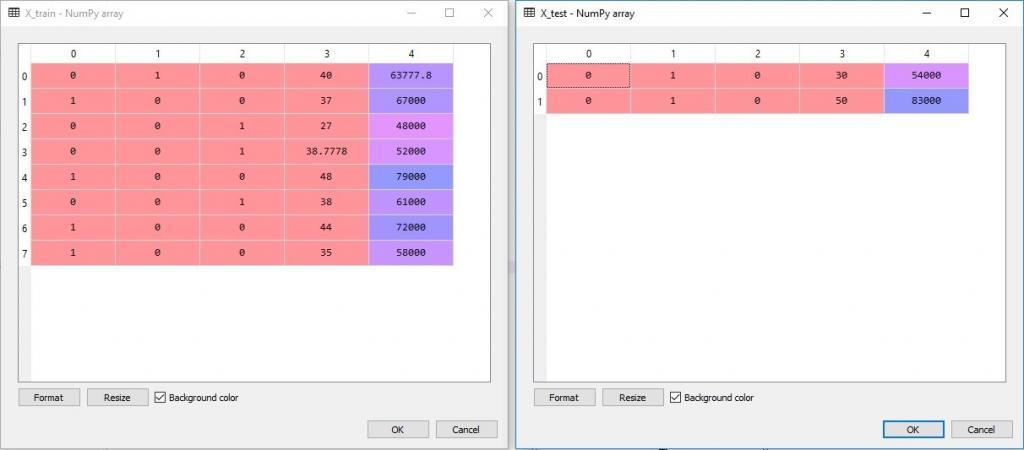
|  |  |
| --- | --- |
| 1  2  3 | # Membagi menjadi training set dan test set  from sklearn.model\_selection import train\_test\_split  X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 0) |

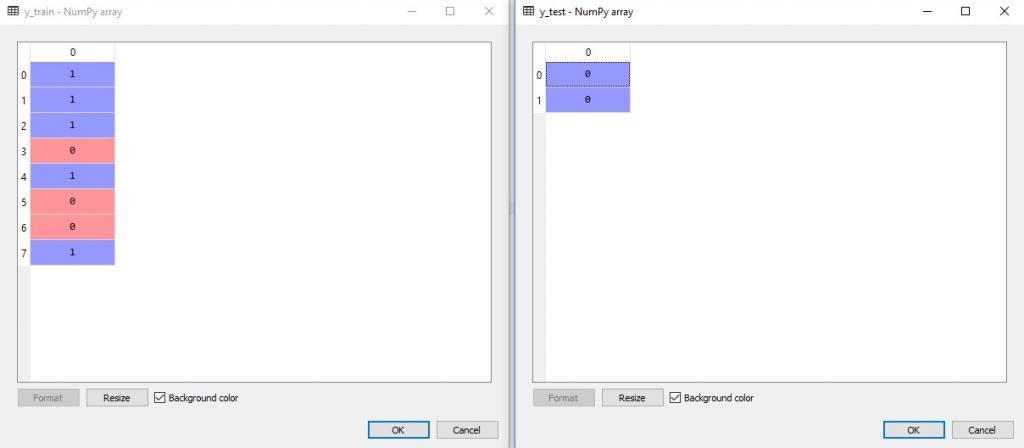
Penjelasan:

Line ke 2 adalah proses import train\_test\_split. Library train\_test\_split ini adalah sublibrary dari modul model\_selection di library sklearn, sehingga penulisan untuk mengimportnya seperti itu (from.. import..)

Line ke 3 adalah proses pendefinisian X\_train sebagai data X yang ingin kita latih (train), X\_test sebagai data X yang ingin kita tes, y\_train adalah variabel dependen yang dilatih, dan y\_test merupakan variabel dependen yang diuji. Kita mendefinisikan keempat variabel ini secara bersamaan. Pada perintah train\_test\_split diisi dengan parameter arrays X,y yang menunjukkan arrays yang dipakai adalah X sebagai variabel independen dan y sebagai variabel dependen. Parameter selanjutnya adalah test\_size sebesar 20%. Biasanya sebagai rule of thumb (aturan umum) proporsi test set sebesar 20% dan train set80%. Namun ada juga yang menentukan 25:75, terserah Anda ingin yang mana. Parameter lainnya adalah random\_state yang merupakan random number generator (RNG). Bisa diisi apa saja, idenya adalah jika menggunakan angka 7 misalnya, maka bilangan random yang dibangkitkan akan selalu sama jika suatu saat kita menentukan angka 7 lagi. Untuk kali ini kita pilih angka nol.

Jika sudah, coba klik dua kali pada X\_train dan X\_test di Variable explorer. Tampilannya akan tmpak sebagai berikut:





Jika kita eksekusi lagi line 3 dengan angka random\_state 7 misalnya, maka nilai X\_test, X\_train, y\_test dan y\_train akan berubah. Namun jika kita eksekusi lagi dengan random\_state nol, maka nilainya akan sama lagi. itulah inti dari random\_state di mana kita akan mendapatkan bilangan random yang sama untuk angka RNG yang sama.

Perlu diingat bahwa logika yang ingin kita bangun adalah kita membiarkan mesin belajar dari training set untuk kemudian diuji di test set. Jadi pada training set, ia akan mencari korelasi antara X\_train dan y\_train. Setelah korelasi ini ada dan terbentuk, kita akan mengujinya ke test set. Jika mesin ini belajar terlalu keras ia akan melakukan overfitting yang akan menghasilkan korelasinya terlau sempurna atau bahkan tidak terbentuk sama sekali. Ada banyak teknik untuk menghindari overfitting dan kita akan membahas ini di pembahasan regresi.

**Analisis masalah**

Lirik lagu mengandung unsur positif dan negatif. **Unsur positif lagu (sentiment positif) yang dimaksud adalah ....** lirik lagu yang mengandung perasaan senang atau gembira

**Unsur negatif lagu (sentiment negatif) yang dimaksud adalah ....** lirik lagu yang mengandung perasaan sedih, iri hati, dendam dll

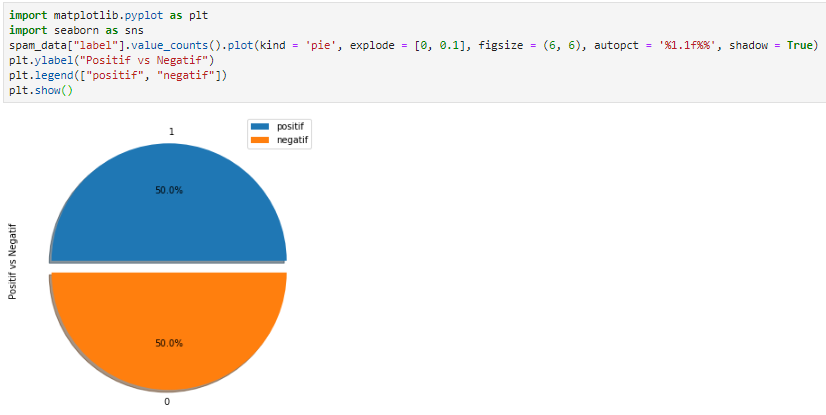
Pembahasan :

1. **Tampilkan daftar lagu**





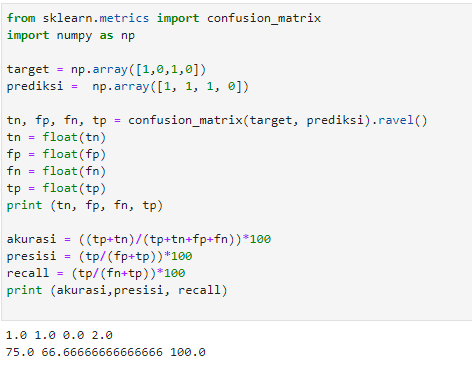
1. **Tampilkan pie chart banyaknya lagu positif dan negatif**



1. **Jelaskan berapa data training dan berapa data testnya**



1. **Tampilkan confusion matrik hasil prediksi / klasifikasi naïve bayes**



**Kesimpulan :**

Metode Naive Bayes dapat digunakan untuk melakukan klasifikasi analisa dengan cukup baik pada sistem analisa sentimen