

LAPORAN IMPLEMENTASI BAYES DECISION RULE

Fauzan Firdaus – 23520011

Link Video : <https://youtu.be/v2W4W1Yc4ZY>

Link Kode, Dataset, Hasil prediksi:

Dalam menyelesaikan tugas ini, saya menggunakan bahasa python dengan tools jupyter notebook. Pertama-tama saya lakukan import library pandas untuk melakukan load dataset. (Penjelasan dataset ada di video). Dalam menyusun programnya, saya buat fungsi/prosedur yang isinya return valuenya adalah data/array yang dihasilkan oleh suatu intruksi.

Fungsi yang pertama ada **countFish**. Fungsi ini bertujuan untuk menghitung jumlah ikan berdasarkan jenisnya, sedangkan parameter masukannya adalah data yang menunjukkan satu kolom dataset berupa kolom *csalmon* atau *cseabass* (count masing-masing jenis ikan).

Selanjutnya ada fungsi **probXgivenY** yang bertujuan untuk mencari peluang lightness given suatu kelas. Fungsi ini melakukan perulangan sebanyak record pada dataset dan menghitung peluangnya pada setiap record tersebut. Syntax yang dilakukan sama dengan rumus peluangnya, yaitu jumlah suatu jenis ikan pada record ke-i dibagi dengan total seluruh jenis ikan tersebut.

Fungsi berikutnya adalah **probYgivenX** yang mendapatkan sebuah nilai posterior conditional probability pada setiap record dalam dataset. Nilai peluang ini dijadikan sebagai penentu decision untuk pendekatan maximum posterior probability. Rumusnya adalah nilai peluang yang didapatkan pada fungsi probXgivenY dikalikan dengan peluang kelasnya saja.

Lalu ada fungsi **decision**. Fungsi ini bertujuan untuk mendapatkan hasil prediksi/decision dari setiap record berdasarkan nilai-nilai peluang yang telah didapatkan. Dalam fungsi ini terdapat 2 kondisi prediksi, pertama jika stat=1, artinya program akan melakukan decision menggunakan pendekatan maximum posterior probability, sedangkan jika stat=2 program akan melakukan decision menggunakan minimum risk. Pada stat=1, perbandingan yang dilakukan adalah membandingkan nilai peluang hasil dari fungsi probYgivenX kelas salmon dengan kelas seabass, nilai peluang yang lebih besar maka kelas tersebutlah yang dipilih sebagai decision. Namun hal ini akan menjadi error pada record pertama datasetnya, karena kebetulan count dari salmon dan seabass sama-sama 0, maka nilai peluangnya 0 juga. Jadi ditambahkan kondisi khusus *else* yang isinya untuk meng-auto generate hasil decision menjadi label 'unknown'. Lalu pada stat=2, yang dibandingkan adalah nilai likelihood dengan threshold yang sudah ditentukan. Jika likelihood < ratio maka akan dikategorikan sebagai seabass, sedangkan sebaliknya akan dikategorikan sebagai salmon.

Fungsi yang terakhir adalah **LikelihoodRatio**. Fungsi ini membagi nilai peluang lightness given kelas salmon dibagi dengan kelas seabass. Untuk kasus mencari likelihood ini, terdapat beberapa kondisi yang menyebabkan error program atau salah decision. Hal ini dikarenakan terjadi pembagian 0 dengan 0 dan pembagian suatu angka dengan 0. Baris pada record seperti pada baris ke 2,3 dst yang memiliki nilai peluang kelas seabass given lightness nya 0 seharusnya dikategorikan sebagai salmon, namun karena hasil likelihoodnya infinite (karena pembagian suatu angka dengan 0) maka terjadi error program. Hal ini dapat disiasati dengan mengganti nilai peluang 0 tersebut menjadi angka yang sangat kecil, sehingga tidak terjadi pembagian yang hasilnya infinite atau NaN.

Threshold yang digunakan pada pendekatan minimum risk ini adalah 0,77 yang didapatkan dari rumus $\frac{\lambda_{12} - \lambda_{22} \frac{P(\omega_2)}{\lambda_{21} - \lambda_{11} \frac{P(\omega_2)}{P(\omega_1)}}}{\lambda_{21} - \lambda_{11} \frac{P(\omega_2)}{P(\omega_1)}}$, nilai λ_{12} dan λ_{21} sama-sama 1 yang menunjukkan nilai alpha salmon dikategorikan sebagai seabass dan nilai alpha seabass dikategorikan salmon. Sedangkan λ_{22} dan λ_{11} sama-sama 0 untuk keadaan sebaliknya.

Program dijalankan dengan memanggil fungsi-fungsi yang telah didefinisikan. Untuk lebih lengkapnya, mohon meninjau kode program dan video presentasi pada link yang tertera.