Nama : Talitha Farah K

Kelas : IF-40-02

Nim : 1301164262

**Laporan Tugas 1 Machine Learning**

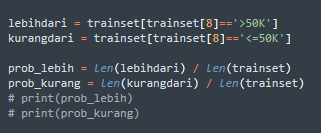
1. **Deskripsi Masalah**

Diberikan sebuah Trainset berupa himpunan data berisi 160 objek data yang memiliki 7 atribut input (age, workclass, education, marital-status, occupation, relationship, hours-per-week) dan 1 output (label kelas income) yang memiliki 4 kelas/label (0, 1, 2, dan 3). Bangunlah sebuah sistem klasifikasi menggunakan metode Naïve Bayes untuk menentukan kelas/label data testing dalam Testset. Sistem membaca masukan file TrainsetTugas1ML.csv dan TestsetTugas1ML.csv dan mengeluarkan output berupa file TebakanTugas1ML.csv berupa satu kolom berisi 40 baris yang menyatakan kelas/label baris yang bersesuaian pada file TestsetTugas1ML.csv.

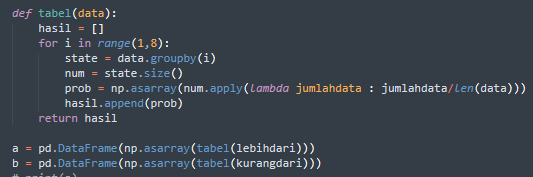
1. Analisis dan Penyelesaian Masalah

Naive Bayes merupakan sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik dengan cara memprediksi peluang di **Data Test** berdasarkan pengalaman di **Data Train**. Tahapan dalam Algoritma ini yaitu :

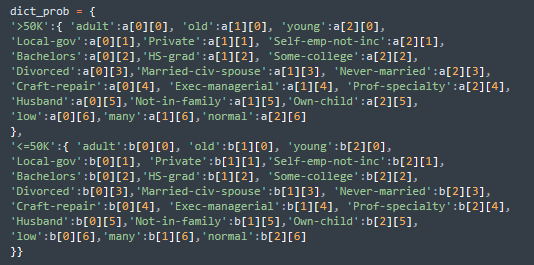
1. Load Data Train
2. Menghitung Probabilitas Keseluruhan

Data Train dibagi menjadi 2 sesuai dengan income nya yaitu “>50K” dan “<=50K”. Variabel prob\_lebih menampung probabilitas keseluruhan dengan income >50K, didapat dari jumlah data dengan income >50K dibagi dengan total data Train. Variabel prob\_kurang menampung probabilitas keseluruhan dengan income <=50K, didapat dari jumlah data dengan income >50K dibagi dengan total data Train.

1. Menghitung Probabilitas Kasus Per Kelas

fungsi tabel berisikan cara menghitung probabilitas data perkelas. data dihitung berdasakan header nya. Kemudian dimasukkan rumus jumlah data/ panjang data menurut kelas nya. Hasil kemudian dimasukkan ke array. Fungsi ini akan menghasilkan probabilitas dari data kurang dari maupun lebih dari.

1. Load Data Test
2. Dictionary Probabilitas

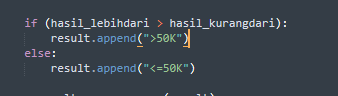
Dictionary ini berisikan probabilitas tiap tiap atribut yang telah dihitung dari fungsi sebelumnya. Dictionary ini berfungsi untuk membandingkan data dari data tes untuk dicari propabilitasnya.

1. Mencari Probabilitas dari Data Tes

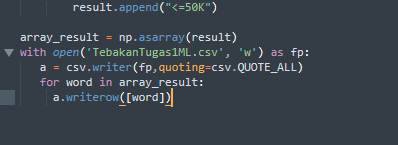
Result berfungsi untuk menampung hasi akhir dari hasil perhitungan. Data Test di load dan ditentukan probabiliasnya. Di cek terus dari atribut ke 1 hingga atribut ke 6. Jika sudah, semua atribut dikalikan dan dikalikan juga dengan probabilitas keseluruhan nya. Hasil kemudian ditampung di variabel hasil\_lebihdari dan hasil\_kurangdari



1. Bandingkan Hasil

Hasil dari keduanya dibandingkan siapa yang lebih besar. Jika hasil\_lebihdari lebih besar dari hasil\_kurangdari maka akan dimasukan “>50K” jika tidak “<=50K”

1. Export Hasil

Hasil kemudian di Export ke TebakanTugas1ML.csv