**LAPORAN TUGAS PROGRAM 1.1**

**Pembelajaran Mesin**

***(Machine Learning)***

**K-Nearest Neighbors Method**

****

Nama : Indah Ayu Nur Fauziah

NIM : 1301164004

Kelas : IF 40-02

**JURUSAN INFORMATIKA**

**FAKULTAS INFORMATIKA**

**UNIVERSITAS TELKOM**

**BANDUNG**

**2018**

1. **Deskripsi Khusus**

Diberikan file TrainsetTugas1ML.csv berupa himpunan data berisi 160 objek 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). Diperlukan sistem untuk mengklasifikasi kelas/label data testisng menggunakan metode Naïve Bayes dalam file TebakanTugas1ML.csv. Sistem membaca masukan file DataTrain\_Tugas3\_AI.csv dan TestsetTugas1ML.csv dan mengeluarkan output berupa file TebakanTugas1ML.csv berupa satu kolom berisi 40 baris yang menyatakan kelas/label baris atau record yang bersesuaian pada file TestsetTugas1ML.csv.

1. **Rancangan Metode**

Permasalahan ini akan diselesaikan menggunakan metode Naïve Bayes*.* Untuk mengklasifikasi kelas pada data testing dalam file TestsetTugas1ML.csv. Algoritma Naïve Bayes adalah metoda klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dr Naïve Bayes Classifier ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Tahapan dari proses algoritma Naive Bayes adalah:

1. Menghitung jumlah kelas / label.
2. Menghitung Jumlah Kasus Per Kelas
3. Kalikan Semua Variable Kelas
4. Bandingkan Hasil Per Kelas

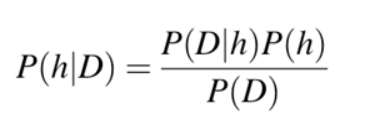


Figure 1. Rumus kelas independensi kondisional

Dimana,

P(h) : probabilitas hipotesis h menjadi benar (terlepas dari data). Dikenal sebagai probabilitas sebelumnya dari h.

P(D) : probabilitas data (terlepas dari hipotesis)

P(h|D) : probabilitas hipotesis data D (probabilitas posterior)

P(D|h) : probabilias data d mengingat bahwa hipotesis h adalah benar.

Rumus diatas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga evidence). Karena itu, rumus diatas dapat pula ditulis sebagai berikut:

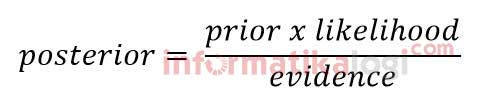


Figure 2. rumus posterior

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

Digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing masing petunjuk saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

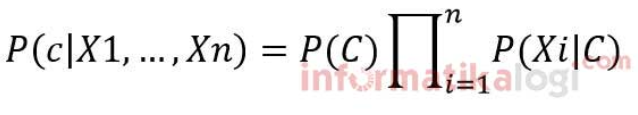


Figure 3. Teorema Naive Bayes

Persamaan diatas merupakan model dari Teorema Naive Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss:

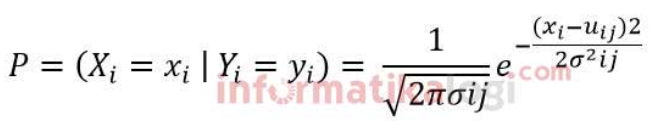


Figure 4. rumus densitas gauss

Keterangan:

P : Peluang

Xi : Atribut ke i

Xi : Nilai atribut ke i

Y : Kelas yang dicari

Yj : Sub kelas Y yang dicari

u : Mean, menyatakan rata rata dari seluruh atribut

o : Deviasi standar, menyatakan varian dari seluruh atribut

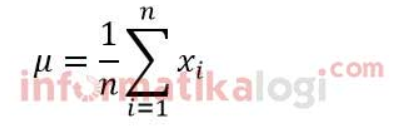


Figure 5. rumus mean

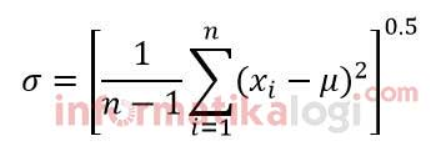
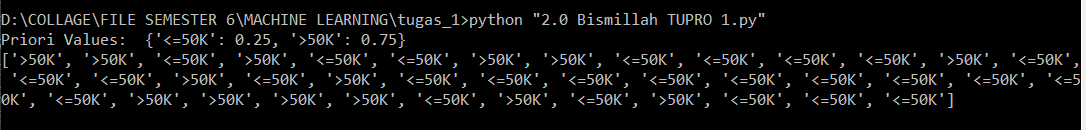


Figure 6. rumus deviasi standar

1. **Screenshoot Program**



Hasil running di cmd:



Hasil file tebakan di csv:

