

Project-Based Learning – Naive Bayes

Nama: Dawnie Julian Nugroho

NIM: 2211104064

Nama: Mohammad Fathurrohman

NIM: 2211104070

1. Deskripsi Kasus

Pada proyek ini, kami membangun sistem sederhana untuk memprediksi kelulusan tepat waktu mahasiswa berdasarkan tiga indikator utama:

- Indeks Prestasi Kumulatif (IPK)
- Persentase Kehadiran Kuliah
- Jumlah Mata Kuliah yang Tidak Lulus

Tujuan utama dari sistem ini adalah untuk mengklasifikasikan apakah seorang mahasiswa akan lulus tepat waktu atau tidak, berdasarkan data akademik historis mahasiswa lainnya.

2. Proses Preprocessing

Karena dataset bersifat kecil dan terstruktur, preprocessing dilakukan secara manual, meliputi:

Normalisasi data tidak dilakukan karena algoritma Naive Bayes Gaussian sudah mempertimbangkan skala melalui varians.

Label dikonversi ke dua kelas: Ya (Lulus Tepat Waktu) dan Tidak.

Data tidak mengandung null value, sehingga tidak dilakukan pembersihan lebih lanjut.

Dataset diisi secara manual dengan asumsi realistis.

Contoh data:

| IPK | Kehadiran (%) | MK Tidak Lulus | Lulus Tepat Waktu |
|-----|---------------|----------------|-------------------|
| 3.2 | 90 | 0 | Ya |
| 2.4 | 75 | 2 | Tidak |
| 3.5 | 95 | 0 | Ya |
| 2.8 | 60 | 3 | Tidak |
| 3.7 | 85 | 0 | Ya |
| 2.5 | 70 | 2 | Tidak |
| 3.0 | 80 | 1 | Ya |
| 2.2 | 65 | 3 | Tidak |

3. Desain Algoritma

Kami menggunakan algoritma Naive Bayes Gaussian, yang menghitung:

- a. Probabilitas prior dari setiap kelas berdasarkan frekuensi data pelatihan.
- b. Distribusi Gaussian (mean dan varians) untuk setiap fitur pada setiap kelas.
- c. Probabilitas gabungan dari fitur input terhadap masing-masing kelas.

Kami tidak menggunakan library seperti sklearn, dan seluruh proses komputasi dilakukan manual dalam kode Python, sesuai instruksi.

4. Evaluasi Model

- a. Model diuji dengan dua data uji.
- b. Prediksi yang diberikan sesuai ekspektasi berdasarkan pola data training.
- c. Tidak dilakukan pembagian data training/testing secara statistik karena dataset kecil.
- d. Evaluasi bersifat logis dan deskriptif.

5. Hasil Program

```
mac@M1-Julian KPL % /usr/local/bin/python3 /Users/mac/Document/KPL/ProjectBasedLearning.py
=== Hasil Prediksi ===
Data 1: [3.1, 88, 0] => Prediksi: Ya
Data 2: [2.3, 60, 3] => Prediksi: Tidak
```

6. Output penjelasan

Program melakukan klasifikasi terhadap dua data baru berdasarkan model Naive Bayes yang dibentuk dari data training. Berikut detail penjelasannya:

Data 1: [3.1, 88, 0]

IPK tinggi (3.1)

Kehadiran tinggi (88%)

Tidak ada MK yang tidak lulus

Berdasarkan data training, kombinasi seperti ini umumnya dimiliki oleh mahasiswa yang lulus tepat waktu. Maka, algoritma Naive Bayes menghitung probabilitasnya lebih tinggi untuk kelas "Ya", dan memberikan prediksi:

Prediksi: Ya

Data 2: [2.3, 60, 3]

IPK rendah (2.3)

Kehadiran rendah (60%)

3 MK tidak lulus

Fitur-fitur ini cocok dengan pola mahasiswa yang tidak lulus tepat waktu dalam data training. Maka, probabilitas untuk kelas "Tidak" lebih tinggi, sehingga hasil prediksinya adalah:

Prediksi: Tidak

Dawnie Julian Nugroho: Perancangan kasus, Implementasi kode Python, dan pengujian output.

Mohammad Fathurrohman: penulisan laporan, dan presentasi, dan dokumentasi kode.