

Implementasi Model Machine Learning untuk Memprediksi Kinerja Akademik Mahasiswa Berdasarkan Tingkat Stres.

Anggota Kelompok:

M. Ilyas Zaini	220411100181
M. Aldi Rahmandika	230411100040
Muhammad Abid Ayyasy	230411100042
Dani Subianto	230411100045
M. Sultan Abdurrahman Al Zudas	230411100191

Data Set

- **Tipe:** Teks (dalam format CSV - Comma Separated Values).
- **Sumber:** Kaggle - Student Stress Monitoring Datasets.
- **Link:**
<https://www.kaggle.com/datasets/mdsultanulislamovi/student-stress-monitoring-datasets/data?select=StressLevelDataset.csv>
- **Detail:** Dataset berisi 1100 data mahasiswa dengan 20 fitur (variabel stres) dan 1 target (kinerja akademik). Fitur dikelompokkan ke dalam kategori psikologis, fisiologis, lingkungan, akademik, dan sosial.

Business Understanding

Tingkat stres mahasiswa memiliki korelasi kuat dengan keberhasilan akademik mereka. Institusi pendidikan perlu memahami faktor-faktor stres yang paling berdampak untuk merancang program dukungan yang efektif. Proyek ini bertujuan untuk mengidentifikasi mahasiswa yang berisiko mengalami penurunan prestasi akibat stres, sehingga intervensi dapat dilakukan lebih dini.

Target/Goal

Membangun sebuah model klasifikasi *machine learning* yang mampu memprediksi kinerja

akademik (*academic_performance*) mahasiswa berdasarkan berbagai faktor stres yang dilaporkan. Tujuannya adalah untuk:

1. Mengidentifikasi faktor stres paling signifikan.
2. Menghasilkan model prediksi dengan akurasi yang andal.
3. Memberikan wawasan untuk mendukung kesejahteraan dan keberhasilan akademik mahasiswa.

Preprocessing

Tahap pra-pemrosesan data akan mencakup langkah-langkah berikut:

1. **Data Cleaning:** Memeriksa dan menangani nilai yang hilang (*missing values*) atau data yang tidak konsisten.
2. **Analisis Fitur:** Menganalisis tipe data dan distribusi setiap fitur.
3. **Pemisahan Data:** Memisahkan antara fitur (variabel stres) dan target (*academic performance*).
4. **Pembagian Dataset:** Membagi data menjadi 80% data latih (*training set*) dan 20% data uji (*testing set*).

Model Building (1 Metode)

- **Metode yang Dipilih: Random Forest Classifier.**
- **Alasan Pemilihan:**
 - Memiliki performa dan akurasi yang tinggi.
 - Kuat (robust) terhadap *overfitting*.
 - Dapat memberikan informasi mengenai fitur mana yang paling penting (*feature importance*), sesuai dengan tujuan proyek.

Pengujian

Kinerja model akan dievaluasi pada data uji menggunakan metrik-metrik berikut:

1. **Confusion Matrix:** Untuk melihat detail performa klasifikasi pada setiap kategori kinerja.
2. **Accuracy:** Mengukur persentase total prediksi yang benar.
3. **Precision, Recall, dan F1-Score:** Untuk memberikan gambaran detail tentang seberapa baik model mengidentifikasi setiap kategori, terutama jika distribusi kelas tidak seimbang.