



KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI  
UNIVERSITAS ANDALAS  
FAKULTAS TEKNOLOGI INFORMASI  
DEPARTEMEN SISTEM INFORMASI  
**KELOMPOK KEILMUAN REKAYASA DATA DAN BUSINESS INTELLIGENCE**  
Kampus Universitas Andalas, Limau Manis, Padang - 25163

---

**INSTRUKSI MODUL 4:**  
**SUPPORT VECTOR MACHINE**

*Dataset* yang kita digunakan adalah *dataset* [Predicting a Pulsar Star](#). Pulsar adalah jenis langka dari bintang Neutron yang menghasilkan emisi radio yang dapat dideteksi di Bumi. Mereka memiliki nilai ilmiah yang tinggi sebagai alat untuk meneliti ruang-waktu, medium antar-bintang, dan keadaan materi. *Dataset* terdiri dari 8 fitur dan 1 kelas. Delapan variabel yang terdapat dalam *dataset* terbagi menjadi dua kelompok:

1. Empat variabel pertama adalah statistik sederhana yang diperoleh dari profil pulsa terintegrasi (Integrated Pulse Profile).
2. Empat variabel lainnya diperoleh dengan cara yang serupa dari kurva DM-SNR (Dispersion Measure - Signal to Noise Ratio).

**Instruksi:**

1. *Import library* dan *dataset* yang akan digunakan.
2. Lakukan Exploratory Data Analysis untuk mendapatkan wawasan mengenai data. Beberapa hal yang disarankan adalah mengubah nama kolom, melihat *summary* data, meng-*handle missing values*, melihat ringkasan statistik (numerikal), memeriksa distribusi kelas target, dan memeriksa *outliers*.
3. Jika data sudah siap untuk digunakan, deklarasikan Vector Feature dan Target Variable-nya. Lalu bagi *dataset* menjadi *data training* dan *data test*.
4. Lakukan Feature Scaling agar semua fitur memiliki skala yang sama.
5. Coba jalankan model dan tampilkan akurasinya untuk:
  - a. *default hyperparameters* (`SVC()`)
  - b. rbf kernel dan  $C=100.0$  (`SVC(C=100.0)`)
  - c. rbf kernel dan  $C=1000.0$  (`SVC(C=1000.0)`)

Berikan kesimpulan dari hasil akurasi masing-masing parameter.



KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI  
UNIVERSITAS ANDALAS  
FAKULTAS TEKNOLOGI INFORMASI  
DEPARTEMEN SISTEM INFORMASI  
**KELOMPOK KEILMUAN REKAYASA DATA DAN BUSINESS INTELLIGENCE**  
Kampus Universitas Andalas, Limau Manis, Padang - 25163

---

(*Opsional*: Coba lakukan dan periksa akurasi untuk jenis kernel yang lainnya, seperti linear kernel, polynomial kernel, dan sigmoid kernel, dengan masing-masing besaran  $C$ , yaitu  $C=1.0$ ,  $C=100.0$ , dan  $C=1000.0$ . Lalu bandingkan akurasi dan berikan kesimpulan).

6. Lakukan Confusion Matrix beserta visualisasinya dan Classification Report. Berikan kesimpulan dari hasil yang didapatkan.
7. Terakhir, cari *hyperparameter* yang paling optimal menggunakan GridSearch CV, dan berikan kesimpulan dari semua hasil analisis yang sudah dilakukan.

**Note:**

Segala bentuk plagiarisme tidak akan pernah dibenarkan. Lakukan instruksi untuk belajar.