

## 1. Topik/Judul

Memprediksi Tingkat Produksi Padi di Sumatera Berdasarkan Faktor Iklim dengan Metode K-Nearest Neighbors (KNN)

## 2. Deskripsi Singkat

Produksi padi di Sumatera dipengaruhi oleh berbagai faktor iklim, seperti curah hujan, kelembapan, dan suhu rata-rata. Penelitian ini bertujuan untuk mengelompokkan hasil produksi padi ke dalam beberapa kategori, misalnya, rendah, sedang, atau tinggi, berdasarkan faktor-faktor iklim. Dengan menggunakan metode K-Nearest Neighbors (KNN), penelitian ini diharapkan bisa membantu para petani dan pengambil kebijakan untuk memaksimalkan hasil panen padi, bahkan ketika kondisi iklim berfluktuasi.

## 3. Masalah dan Tujuan

Masalah:

- Bagaimana kita bisa memanfaatkan faktor-faktor iklim untuk memprediksi tingkat produksi padi menggunakan metode K-Nearest Neighbors.
- Apakah faktor-faktor iklim ini cukup akurat dalam mengelompokkan hasil panen padi

Tujuan:

Mengelompokkan tingkat produksi padi menjadi beberapa kelas (rendah, sedang, tinggi) berdasarkan data iklim. Menilai seberapa baik metode K-Nearest Neighbors dalam memprediksi kelompok produksi padi. Memberikan rekomendasi untuk para petani dan pembuat kebijakan tentang bagaimana memaksimalkan produksi padi dengan mempertimbangkan faktor-faktor iklim.

## 4. Alur/Tahapan Penelitian

1. Pengumpulan dan Pemahaman Data:  
Mengumpulkan data produksi padi beserta faktor iklim (curah hujan, kelembapan, dan suhu rata-rata) dari berbagai provinsi di Sumatera. Memahami setiap variabel untuk memastikan data yang relevan dan mengenal keterbatasan dataset.
2. Pra-pemrosesan Data:
  - Menangani nilai yang hilang serta menyetarakan skala data (normalisasi) untuk meningkatkan kinerja algoritma.
  - Mengonversi kolom "Produksi" menjadi kategori (misalnya, rendah, sedang, tinggi) berdasarkan distribusi data.
3. Eksplorasi Data:  
Menganalisis pola umum pada data, termasuk distribusi produksi padi dan variasi kondisi iklim di berbagai provinsi. Menggali hubungan awal antar variabel yang bisa mendukung model prediksi.
4. Penerapan Algoritma:
  - Mengimplementasikan algoritma K-Nearest Neighbors (KNN).

- Menentukan nilai k terbaik dengan melakukan eksperimen, untuk mendapatkan kinerja optimal dalam mengelompokkan tingkat produksi padi.
- 5. Evaluasi Model:  
Mengukur performa model menggunakan metrik evaluasi seperti akurasi, precision, recall, dan skor F1 untuk menilai seberapa baik model memprediksi kategori produksi.
- 6. Visualisasi dan Interpretasi:  
Menyajikan hasil model melalui visualisasi, seperti grafik hubungan faktor iklim dan kategori produksi, serta menafsirkan faktor iklim yang paling berpengaruh dalam model.
- 7. Penerapan:  
Menggunakan model yang telah dilatih untuk memprediksi kategori produksi padi berdasarkan data iklim baru, serta memberikan rekomendasi praktis bagi petani dan pembuat kebijakan di Sumatera untuk memaksimalkan hasil panen.

## 5. Penjelasan Dataset

1. Sumber Data:  
Dataset ini berasal dari file "Data\_Tanaman\_Padi\_Sumatera\_version\_1.csv" yang berisi informasi tentang produksi padi, luas panen, dan data iklim untuk berbagai provinsi di Sumatera. [Dataset Tanaman Padi Sumatera, Indonesia](#)
2. Penjelasan Kolom Data:
  - Provinsi: Nama provinsi di Sumatera.
  - Tahun: Tahun pengambilan data.
  - Produksi: Jumlah hasil produksi padi, yang akan dikelompokkan menjadi kategori (rendah, sedang, tinggi).
  - Luas Panen: Luas lahan panen (hektar).
  - Curah Hujan: Curah hujan tahunan (dalam mm).
  - Kelembapan: Tingkat kelembapan rata-rata (%).
  - Suhu Rata-rata: Suhu rata-rata tahunan (dalam derajat Celsius).