

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Tanaman padi atau *Oryza sativa L.* merupakan tanaman budidaya yang sangat penting karena beras menjadi sumber makanan pokok bagi sebagian besar penduduk Asia. Tanaman ini tumbuh di daerah beriklim tropis dan subtropis yang memiliki suhu panas dan lembab, seperti di Indonesia. Kondisi iklim di Indonesia sangat mendukung pertumbuhan padi, sehingga menjadikan negara tersebut sangat cocok untuk budidaya padi.

Provinsi Jawa Timur adalah provinsi yang menghasilkan padi terbanyak di Indonesia, sehingga provinsi tersebut dituntut agar produksi padi terus meningkat guna mencapai kebutuhan pangan. Baik untuk memenuhi kebutuhan masyarakat Jawa Timur maupun masyarakat secara umum. Kabupaten Jember termasuk sentra produksi padi di Jawa Timur dengan produksi padi pada tahun 2022 sebanyak 607,37 ribu ton GKG (gabah giling kering) dan memiliki luas panen mencapai 118,49 ribu hektar. Pada tahun 2023, produksi padi meningkat menjadi sekitar 616,73 ribu ton GKG dengan luas panen mencapai sekitar 120,19 ribu hektar (BPS Kabupaten Jember, 2023). Kabupaten Jember secara administratif memiliki 31 kecamatan dengan 226 desa dan 22 kelurahan.

Pengelolaan data produksi padi perlu dioptimalkan dengan menerapkan metode *clustering* dan visualisasi untuk pengolahan serta penyimpanan data dalam jumlah besar. Berdasarkan observasi di BPS Kabupaten Jember, saat ini belum tersedia sistem yang mengelompokkan data produksi padi pada tingkat desa, hanya tersedia pada tingkat kecamatan atau kabupaten. Dengan memanfaatkan *data mining*, pola-pola baru berdasarkan nilai produksi padi setiap desa dapat ditemukan. Untuk itu, perlu sebuah sistem yang mampu mengelompokkan desa-desa di Kabupaten Jember berdasarkan kategori tingkat tinggi, sedang maupun rendah. Selain itu, visualisasi data menggunakan peta geografis sangat diperlukan para petani maupun masyarakat agar dapat melihat data tersebut berupa luas tanam, luas panen, dan produksi padi di setiap desa.

Sistem Informasi Geografis merupakan sistem komputer yang dirancang sebagai media penyimpanan, pengelolaan, analisis, dan penyebaran informasi geografis yang berkaitan dengan tata ruang serta data yang berhubungan dengan permukaan bumi (Prasetya dkk., 2021). Dalam mengolah data besar, data mining digunakan untuk mengidentifikasi pola, hubungan, dan trend tersembunyi agar dapat membantu dalam mengambil keputusan. Proses ini meliputi pengumpulan, pengolahan, dan analisis data menggunakan metode statistik, matematika, serta teknologi seperti kecerdasan buatan.

Dalam *data mining*, *clustering* digunakan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kemiripan jenis data. Terdapat beberapa algoritma dalam *clustering*, salah satunya adalah *K-Means Clustering* (Wijayanto & Yoka Fathoni, 2021). Algoritma ini merupakan algoritma pembelajaran sederhana yang mampu menyelesaikan masalah dengan mengurangi kesalahan tersebut.

Beberapa penelitian terdahulu yang menggunakan metode *K-Means Clustering* dan *DBI* diantaranya adalah sebagai berikut. Penelitian yang dilakukan oleh (Wahidah dkk., 2023) membahas tentang pengelompokan daerah rawan bencana di Kabupaten Jember menggunakan metode *K-Means Clustering*. Penelitian ini mengelompokkan kecamatan ke dalam beberapa kelompok berdasarkan kategori tingkat kerawanan tinggi, sedang, dan rendah dengan menghasilkan jumlah *cluster* yang paling optimal adalah 3 dan nilai *DBI*-nya sebesar 1,156763. Penelitian oleh (Ananda dkk., 2022) tentang implementasi *K-Means* dalam pengelompokan data akta kelahiran di Indonesia. Penelitian ini menghasilkan nilai *DBI* sebesar 0,059 dengan jumlah *cluster* yang paling optimal adalah 4. Selanjutnya, penelitian yang dilakukan oleh (W. Putri & Afdal, 2023) membahas tentang penerapan algoritma *K-Means* untuk pengelompokan data penyandang disabilitas di Kabupaten Roka Hilir. Penelitian ini mengelompokkan kecamatan yang menghasilkan 3 *cluster* berdasarkan kategori tingkat disabilitas rendah, sedang, dan tinggi dengan nilai *DBI* yang dihasilkan adalah 0,063.

Berdasarkan penelitian-penelitian sebelumnya, algoritma *K-Means Clustering* dipilih dalam penelitian ini karena mudah untuk dipahami dan

diimplementasikan, serta efektif dalam mengelompokkan data numerik dalam jumlah besar berdasarkan kemiripan nilai seperti mengelompokkan luas tanam, luas panen, dan produksi padi setiap desa di Kabupaten Jember. Untuk memastikan bahwa hasil *clustering* yang diperoleh menunjukkan jumlah cluster yang optimal, penelitian ini menggunakan *Davies Bouldin Index (DBI)* yang digunakan untuk mengevaluasi kualitas *clustering*. Semakin kecil nilai *DBI*-nya, maka kualitas *clustering* yang dihasilkan semakin baik.

Berdasarkan latar belakang diatas, maka penelitian yang akan dilakukan adalah mengelompokkan produktivitas panen padi setiap desa di Kabupaten Jember menggunakan *K-Means Clustering* dan *Davies Bouldin Index* dengan tujuan untuk memberikan informasi pemetaan terkait produksi padi yang dapat diakses oleh masyarakat, petani maupun instansi dalam mengambil keputusan yang lebih tepat untuk pengelolaan dan peningkatan produksi padi.

## **1.2 Rumusan Masalah**

Rumusan masalah yang didapat dari permasalahan diatas adalah sebagai berikut:

1. Berapa nilai *Davies Bouldin Index (DBI)* yang optimal untuk menentukan jumlah *cluster* terbaik?
2. Berapa jumlah *cluster* optimal yang terbentuk untuk membangun Sistem Informasi Geografis (SIG)?

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah:

1. Menentukan nilai *DBI* yang optimal dan memilih jumlah *cluster* terbaik.
2. Menentukan jumlah *cluster* yang paling optimal untuk membangun Sistem Informasi Geografis (SIG).

## **1.4 Manfaat Penelitian**

Adapun manfaat yang diperoleh dari penelitian ini adalah:

1. Dapat membantu BPS Kabupaten Jember dalam menyediakan visualisasi data produksi padi pada tingkat desa melalui pengelompokan data

menggunakan metode *K-Means Clustering* dan hasilnya divisualisasikan dalam bentuk peta digital berbasis Sistem Informasi Geografis. Agar masyarakat maupun petani dapat mengetahui secara jelas produksi padi setiap desa di Kabupaten Jember.

2. Dapat dijadikan sebagai referensi bagi penelitian selanjutnya menggunakan metode *K-Means Clustering* dan *Davies Bouldin Index (DBI)*.

### 1.5 Batasan Masalah

Ditetapkan batasan-batasan masalah agar tidak menyimpang dari tujuan utama, antara lain:

1. Data dalam penelitian ini bersumber dari publikasi resmi BPS Kabupaten Jember pada tahun 2022 dan 2023.
2. Penelitian ini menggunakan 248 data yang mencakup seluruh desa di Kabupaten Jember.
3. Pada penelitian ini, menggunakan variabel berupa luas tanam, luas panen, dan produksi padi.
4. Penelitian ini menggunakan algoritma *K-Means Clustering* dan *DBI*.
5. Sistem Informasi Geografis berbasis web untuk visualisasi hasil *clustering* menggunakan *library* Leaflet dan data peta dalam format *GeoJSON*.
6. Sistem dirancang menggunakan *framework Flask Python*.