

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Kemajuan perangkat keras dan perangkat lunak aplikasi yang semakin canggih menjadikan pengolahan data mentah juga semakin kompleks. Pengolahan data mentah yang kompleks menjadikan proses dalam mengekstraksi data menjadi rumit untuk menggali sebuah informasi yang berguna di sebuah data mentah. Pengolahan data yang kompleks tidak sebanding dengan kemampuan manusia dalam mengolah data sehingga pengolahan sebuah data mentah menggunakan sebuah metode untuk menggali sebuah pola hubungan atau keteraturan di sebuah data. Metode pendekatan yang dapat digunakan yaitu data *mining*.

Data *mining* adalah proses mengekstraksi dan mengidentifikasi informasi bermanfaat dan pengetahuan dari berbagai basis data besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* (Pradana, 2019). Salah satu fungsi data *mining* yaitu *clustering*, *clustering* adalah proses dimana sekumpulan data atau objek akan dipisah kedalam kelompok atau *cluster* yang lebih kecil berdasarkan kesamaan ciri yang dimiliki (Putra, 2017). Salah satu jenis *clustering* yang dapat digunakan untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan antar data yaitu metode K-Means. Metode K-Means merupakan salah satu metode yang populer karena mudah dalam implementasi ke berbagai jenis kasus dan atribut data, selain kemudahan implementasi metode K-Means juga masuk dalam 10 besar metode data mining yang paling digunakan (Nishom, 2018)

K-Means *clustering* adalah Metode pengelompokan yang membagi data ke dalam kelompok yang memiliki karakteristik sama akan dikelompokkan kedalam kelompok yang sama sehingga *dataset* tidak tumpang tindih Sirait (2017). Algoritma K-Means konvensional secara umum dalam penentuan jumlah *cluster* dilakukan oleh *user* dan dalam penentuan nilai *centroid* untuk

tiap *cluster* dilakukan secara acak kemudian data akan ditempatkan ke dalam *cluster* berdasarkan kedekatan jarak yang ada (Putra, 2017). Menurut Salsabila (2019) pada metode *clustering* K-Means memiliki kekurangan dalam pemilihan *centroid* awal yang acak dapat membuat hasil yang berbeda-beda, maka disarankan menggunakan metode yang cocok dalam pemilihan *centroid* awal sehingga akan meningkatkan akurasi pada proses *clustering*. Sedangkan menurut Gustientiedina (2019) metode *clustering* K-Means perlu dilakukan penentuan nilai *centroid* terbaik agar proses *clustering* data dapat dilakukan lebih valid.

*Rank Order Centroid* (ROC) merupakan metode yang sederhana, dapat menghasilkan nilai bobot sesuai kriteria yang digunakan, kinerja tertinggi dalam hal identifikasi alternative dan dapat mengatasi pembobotan atribut dalam pengambilan keputusan (Addenan, 2021). Kelebihan ROC adalah dapat memperlihatkan kriteria yang lebih penting atau diprioritaskan sampai akhir kriteria dan dalam menentukan nilai bobot didapat dari urutan tingkat prioritas kriteria dimulai dari urutan pertama, kedua, ketiga dan seterusnya (Panjaitan, 2021).

Penelitian ini menggunakan *dataset seeds* atau biji gandum, karena penggunaan *dataset seeds* bertujuan mengembangkan pengelompokan biji gandum berdasarkan kelompok yang diinginkan menggunakan metode K-Means, sehingga individu atau petani atau perusahaan akan lebih mudah dalam mengelompokkan berbagai biji yang berbeda. Penelitian ini memprioritaskan biji yang memiliki kriteria area biji yang besar untuk prioritas utama dalam pemilihan biji.

Penelitian ini akan memperbaiki kelemahan algoritma K-Means yang penentuan nilai *centroid* awal secara acak, sehingga pada penelitian ini akan membahas penentuan nilai *centroid* awal dengan menggunakan metode *Rank Order Centroid* dan K-Means konvensional dengan studi kasus *seed dataset* atau data biji gandum.

## 1.2 Perumusan Masalah

Bagaimana menggunakan metode *Rank Order Centroid* (ROC) dalam penentuan nilai *centroid* pada algoritma K-Means.

## 1.3 Batasan Masalah

Agar penelitian ini tetap pada fokus utamanya dan tidak terlalu luas cakupannya maka peneliti menerapkan beberapa batasan masalah. Batasan masalah tersebut antara lain yaitu:

1. Jumlah *cluster* pada penelitian ini ada empat yaitu dua *cluster*, tiga *cluster*, empat *cluster*, dan lima *cluster* dengan batas iterasi maksimal sebanyak 10 iterasi.
2. Pengujian kinerja pada penelitian ini adalah pengujian akurasi metode *Rank Order Centroid* pada K-Means menggunakan *Davies-Bouldin Index*.

## 1.4 Tujuan dan Manfaat

### 1.4.1 Tujuan

Penelitian ini memiliki tujuan yaitu menghasilkan peningkatan nilai akurasi algoritma K-Means dengan *centroid* awal ditentukan oleh hasil *Rank Order Centroid* dan membandingkannya dengan K-Means konvensional.

### 1.4.2 Manfaat

Manfaat dari penelitian ini adalah :

1. Memperkaya wawasan dan kemampuan dalam penelitian dalam data *mining* khususnya dalam metode *clustering* di metode K-Means untuk suatu *dataset*.
2. Hasil penelitian ini diharapkan penggunaan dalam metode K-Means dapat lebih efektif dalam proses *clustering* dan meningkatkan nilai akurasi dalam *clustering*.
3. Hasil penelitian ini diharapkan untuk individu atau petani maupun perusahaan dapat dijadikan sebagai model dalam *clustering* yang efektif dan akurat sesuai kriteria yang dihasilkan dari biji gandum.

### 1.5 Sistematika Penulisan

Sistematika Penulisan sebagai berikut :

## **BAB I PENDAHULUAN**

Pada bab I berisi tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan penelitian penggunaan metode *ROC* dalam penentuan nilai *centroid* (studi kasus : *dataset* biji gandum).

## **BAB II LANDASAN TEORI**

Pada bab II berisi tentang penelitian terdahulu, kerangka pemikiran, dan landasan teori seperti data *mining*, *clustering*, algoritma K-Means, *Rank Order Centroid*, dan *Davies-Bouldin Index* (DBI).

## **BAB III METODOLOGI PENELITIAN**

Pada bab III berisi tentang jenis *dataset* yang digunakan dan metode *rank order centroid* dalam penentuan *centroid* awal di metode K-Means.

## **BAB IV HASIL DAN PEMBAHASAN**

Pada bab IV berisi tentang hasil dari penelitian menggunakan metode K-Means konvensional dan metode *Rank Order Centroid* dalam penentuan *centroid* awal di metode K-Means dengan menggunakan *dataset* biji gandum.

## **BAB V SIMPULAN DAN SARAN**

Pada bab V berisi tentang kesimpulan dari penelitian dan saran-saran untuk melengkapi dan memperbaiki penyusunan penelitian.

## **DAFTAR PUSTAKA**

## **LAMPIRAN**