

# **BAB 1**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Indonesia merupakan salah satu negara produksi pertanian Kakao, namun pertumbuhan hasil produksi pertanian Kakao ini mengalami penurunan dengan rata-rata 2,01% dalam sepuluh tahun terakhir dari lima negara teratas penghasil kakao [1]. Sementara itu, permintaan akan produk ini tumbuh 3% per tahun dan lima besar negara penghasil kakao tersebut memproduksi lebih dari 95% kebutuhan kakao dunia. Rendahnya pertumbuhan produksi dan produktivitas kakao pertanian kakao di Indonesia didasarkan beberapa alasan seperti keterbatasan pengetahuan petani dalam budidaya dan pengelolaan tanaman kakao, masalah pemasaran dan lainnya [1]. Oleh karena itu sangat penting untuk memahami tren produksi kakao khususnya pengelompokan data petani dan hasil produksinya sehingga menghasilkan informasi yang akurat yang dapat berguna untuk pengambilan keputusan.

Dalam beberapa tahun terakhir ini, metode data mining merupakan salah satu metode yang banyak digunakan untuk menggali informasi yang berguna dari sekumpulan data salah satunya adalah algoritma cluster seperti K-Means Clustering, dbSCAN clustering, hierarchical clustering, fuzzy c means dan lainnya. Analisis cluster merupakan salah satu algoritma yang digunakan untuk mengelompokkan basis data sehingga data dalam kluster serupa, dan seberbeda mungkin dengan data di kluster lain. Pengelompokan memungkinkan interpretasi mendalam dengan banyak implikasi tentang data mana yang harus ditargetkan dengan data tertentu yang paling mungkin menarik [2]. Algoritma pengelompokan partisi, seperti K-means menetapkan objek ke dalam k (nomor cluster yang telah ditentukan) cluster, dan merealokasi objek secara iteratif untuk meningkatkan kualitas hasil pengelompokan [3].

Algoritma pengelompokan hirarkis menetapkan objek dalam cluster terstruktur pohon, yaitu, sebuah cluster dapat memiliki perwakilan titik data dari cluster tingkat rendah. Ide metode pengelompokan berbasis Kepadatan adalah bahwa untuk setiap titik dari sebuah cluster, lingkungan dari unit jarak tertentu mengandung setidaknya sejumlah titik minimum, yaitu kepadatan di lingkungan tersebut harus mencapai beberapa ambang batas. Ide dari algoritma pengelompokan berbasis kepadatan adalah bahwa, untuk setiap titik dari sebuah cluster, lingkungan dari unit jarak tertentu harus

mengandung setidaknya sejumlah titik minimum. Penerapan analisis cluster banyak diterapkan di berbagai bidang seperti pengelompokan data pelanggan [2], [4]–[6], pemetaan produktivitas panen [7], data pertanian [3], hasil produksi kelapa sawit [8], pengelompokan hasil buah [9] dan lainnya

Dalam makalah ini, kami bertujuan untuk menerapkan algoritma cluster dalam bidang pertanian yang secara khusus diterapkan untuk pemetaan data pertanian kakao di 4 (empat) provinsi di Indonesia yaitu Sumatera Utara, Sumatera Barat dan Aceh. Sumber dataset merupakan sekumpulan data dari hasil survey yang memuat informasi tentang petani, lahan dan hasil produksi. Pendekatan algoritma clustering K-Means Clustering dan Gaussian Mixture Models digunakan untuk mengelompokkan data hasil produksi kakao

## **1.2. Rumusan Masalah**

Hasil produksi pertanian kakao di Indonesia pada saat ini sangat rendah sementara permintaan setiap tahunnya terus meningkat, sehingga sangat penting untuk membangun sebuah model yang dapat mengelompokkan data pertanian kakao. Oleh karena ini, maka dalam penelitian ini yang menjadi rumusan masalah adalah apakah algoritma cluster dapat digunakan menghasilkan pengelompokan data pertanian kakao yang berbeda?

## **1.3. Batasan Masalah**

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma cluster K-Means Clustering dan Gaussian Mixture Models untuk menganalisis data pertanian kakao.
2. Menggunakan bahasa pemrograman python.
3. Menggunakan dataset hasil survey pertanian kakao.

## **1.4. Tujuan Penelitian**

Tujuan utama penelitian ini adalah menganalisis data pertanian menggunakan teknik data mining yang secara khusus menggunakan algoritma clustering K-Means Clustering, K-Means++ dan Gaussian Mixture Models.

### **1.5. Manfaat Penelitian**

Penelitian ini diharapkan memiliki manfaat diantaran:

1. Sebagai Bahan masukan bagi pemerintah untuk meningkatkan hasil produktivitas pertanian kakao
2. Sebagai bahan referensi bagi penelitian yang tertarik dibidang data mining khususnya dengan teknik pengelompokan

### **1.6. Keterbaruan**

Teknik pengelompokan data mining seperti K-Means merupakan salah satu algoritma yang banyak diterapkan oleh banyak peneliti diantaran [10] menggunakan algoritma clustering untuk pengelompokan data pelanggan caffe dengan tujuan untuk mengetahui apakah nilai-nilai seperti perlindungan lingkungan, perilaku etis produsen, perdagangan yang adil atau memaksimalkan fungsi utilitas konsumsi merupakan faktor penting dalam proses pembelian kopi dan jika mereka dapat digunakan sebagai variabel segmentasi, kemudian [3] menerapkan algoritma clustering yang berbeda pada data pertanian di Negara India dan [8] mengusulkan algoritma K-Means untuk memetakan data hasil pertanian kelapa sawit.