

BAB I

PENDAHULUAN

1.1. Latar Belakang

Bencana gempa bumi merupakan salah satu bencana alam yang paling sering terjadi terutama di Indonesia karena berada di tiga lempeng bumi sekaligus, yaitu lempeng Eurasia, Indo-Australia, dan lempeng Pasifik.

Berdasarkan data yang didapatkan data website USGS tercatat kurang lebih 12.343 kali gempa yang terjadi di Indonesia dalam rentang tahun 2000 sampai 2005. Dengan sangat bergamnya pola, nilai, dan karakteristik dari kejadian gempa bumi yang banyak tersebut tentu saja membuat perhitungan dan pengelompokan data akan sulit dilakukan secara manual. Penggunaan algoritma pembelajaran mesin dapat diterapkan untuk memproses data yang memiliki banyak fitur atau karakteristik. *Data mining* merupakan bagian dari pembelajaran komputer yang bertujuan untuk ekstraksi data dari *dataset* dan mengubah data tersebut menjadi sebuah informasi baru yang memiliki struktur dan dapat dipergunakan untuk kebutuhan pengambilan keputusan kedepannya. *Data mining* memiliki banyak metode salah satunya adalah klasterisasi.

Klasterisasi adalah salah satu teknik yang fundamental dalam menganalisis data yang didalamnya dilakukan penggabungan data yang memiliki karakteristik data yang mirip menggunakan sebuah proses *unsupervised*. *Unsupervised learning* merupakan nama lain dari teknik klasterisasi dengan proses utamanya adalah membagi sebuah dataset yang belum memiliki label menjadi sebuah kelompok yang baru. Anggota klaster dinilai berdasarkan kemiripan karakteristik data. Untuk mencapai sebuah kelompok yang memiliki nilai validitas tinggi, perhitungan kemiripan data antar objek data di dalam satu kelompok/klaster harus semirip mungkin dan perhitungan kemiripan antar objek dalam kelompok/ klaster harus memiliki kemiripan yang minim.

Salah satu metode yang dapat digunakan dalam klasterisasi *big data* adalah algoritma BIRCH (*Balance interative reducing and Clustering*

Using Hierarchies). Algoritma BIRCH merupakan algoritma yang sudah ada sejak 1996 yang dikembangkan oleh Zhang, Ramakrishnan, dan Livny. Algoritma ini mempunyai sistem yang disebut dengan *cluster feature (CF)*. *CF (cluster features)* akan menyimpan dan menyatukan informasi yang terdapat pada *-sub-cluster (CF tree)*. Kemampuan dalam mengubah data menjadi *sub-cluster* ini membuat algoritma BIRCH sangat baik untuk pengolahan *big data* dalam waktu singkat [1].

Terdapat banyak penelitian sudah dilakukan menggunakan algoritma BIRCH salah satunya digunakan untuk deteksi anomali trafik [2]. Menurut penelitian ini, nilai akurasi algoritma BIRCH bisa mencapai 98,45% dan nilai akurasi akan semakin tinggi apabila dataset memiliki karakteristik kemiripan yang semakin minim. Namun pada penelitian tersebut, *Threshold* algoritma BIRCH tidak dilakukan *update* dikarenakan akan memakai lebih banyak waktu untuk memproses klaster.

Penggunaan algoritma BIRCH juga telah digunakan pada penelitian untuk pengelompokan data sekolah menengah atas di daerah DIY Yogyakarta [3]. pada penelitian ini klasterisasi siswa dilakukan berdasarkan indeks integritas ujian nasional dan didapatkan hasil berupa tiga klaster data.

Penelitian Tugas Akhir ini melakukan klasterisasi menggunakan algoritma BIRCH dengan dataset gempa bumi yang terjadi di Indonesia untuk mendapatkan klaster data yang dapat dibaca sehingga bisa dipergunakan untuk keperluan pengembangan informasi, pembelajaran, dan pengetahuan. Salah satunya di bidang mitigasi bencana dan lain-lainnya.

1.2. Rumusan Masalah

1. Bagaimana performansi algoritma BIRCH dalam memproses dataset yang tersedia?
2. Berapa jumlah klaster yang paling optimal dalam klasterisasi yang dihasilkan?
3. Fitur gempa apa saja yang paling berperan dalam klasterisasi gempa?

1.3. Tujuan

1. Mencari klasterisasi gempa berdasarkan parameter seismik.
2. Menilai tingkat kerawanan gempa berdasarkan fitur yang paling berpengaruh.
3. sebagai salah satu pengembangan mitigasi bencana.

1.4. Batasan Masalah

Batasan masalah dalam penelitian ini, yaitu :

- a) Menggunakan bahasa pemrograman python.
- b) Metode klasterisasi menggunakan algoritma BIRCH.
- c) Penggunaan library scikit-learn.
- d) Dataset berisikan indeks/atribut yang dibutuhkan dengan format data .csv.