

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Internet telah mengubah bagaimana suatu informasi ataupun berita dapat disebar dengan mudah. Banyak lembaga media yang memanfaatkannya sebagai sarana publikasi. Salah satunya adalah lembaga penyedia berita. Para penerbit surat kabar saat ini berlomba-lomba menyajikan beritanya dalam versi online. Tentunya hal ini memudahkan seseorang dalam memperoleh berita dengan cepat dan mudah (Karhendana, 2008)

Berita elektronik berbahasa Indonesia volumenya pasti semakin hari bertambah besar. Berita yang ditulis dan disebarluaskan melalui media elektronik ini tentu merupakan sumber informasi yang berharga. Oleh karena itu dibutuhkan sebuah metode khusus untuk dapat mengelompokkan berita-berita tersebut sehingga dapat mempermudah pengambilan informasi penting yang ada di dalamnya. *Clustering* dokumen teks adalah salah satu operasi pada *text mining* untuk mengelompokkan dokumen yang memiliki kemiripan isi. *Clustering* dapat diaplikasikan untuk menemukan keterkaitan antar berita. *Clustering* dapat digunakan untuk membantu menganalisis berita dengan mengelompokkan secara otomatis berita yang memiliki kemiripan (*similarity*).

Teknik *clustering* saat ini juga telah banyak digunakan untuk mengatasi permasalahan yang terkait dengan segmentasi data. Implementasi *clustering* ini

dapat diterapkan pada berbagai bidang sebagai contoh dalam hal *text mining*. Teknik *clustering* dapat digunakan sebagai metode dalam mengelompokkan dokumen teks yang memiliki kesamaan konten/isi dan tema dari teks tersebut. Tujuan utama dari metode *cluster* adalah pengelompokan sejumlah data atau obyek ke dalam *cluster* (kelompok) sehingga dalam setiap *cluster* akan berisi data yang memiliki kesamaan karakteristik dari data tersebut. Pada pengelompokan berita, sekumpulan berita yang belum diberi label kelasnya akan dikelompokkan sesuai dengan karakteristik-karakteristik kata yang dimiliki setiap dokumen tersebut. Hal tersebut dilakukan untuk memudahkan pengorganisasian dokumen pada kebutuhan lebih lanjut.

Ada berbagai macam metode *clustering* salah satunya adalah Fuzzy C-Means ditemukan pertama kali oleh Dunn pada tahun 1973 kemudian dikembangkan lagi oleh Bezdek pada tahun 1981. Ide dasar dari metode ini mirip dengan metode K-Means. FCM didasarkan pada logika fuzzy, setiap titik data dimasukkan ke suatu kelompok berdasarkan nilai keanggotaannya pada kelompok tersebut. (Emha Taufiq Luthfi, 2007)

Dalam penelitian ini, data-data berita yang berkaitan dengan ekonomi, kesehatan, kriminal, olahraga dan politik akan dijadikan data latih atau sample, dan dijadikan acuan bagi algoritma adalah judul dan penulis menggunakan Fuzzy C-means untuk membentuk pusat *centroid* dalam menentukan *cluster-cluster* berita nantinya dan Algoritma Fuzzy C-Means akan diterapkan melalui *tools* rapid miner yang telah ditambahkan fitur ekstensinya. Dari uraian di atas maka penulis mengajukan sebuah judul, yaitu “Pemanfaatan *Text Mining* Pada Artikel Berita untuk menentukan tipe berita menggunakan metode Fuzzy C-Means”

dengan harapan agar memaparkan hasil analisis teks berita tersebut dengan bantuan rapid miner.

## **1.2. Perumusan Masalah**

Berdasarkan latar belakang diatas diperoleh suatu rumusan masalah yaitu “Bagaimana Mengimplementasikan Metode Fuzzy C-Means Untuk Mengelompokkan Dokumen Berita menurut tipenya ”.

## **1.3. Batasan Masalah**

Untuk lebih mengarahkan masalah yang ada serta tidak menyimpang dari permasalahan yang akan dilakukan dalam penelitian. Maka, penulis hanya membatasi pada permasalahan berikut ini:

1. Data sampel menggunakan data primer. Data primer diperoleh dari situs berita *online* seperti detik ,kompas dan cnnindonesia.
2. Kategori berita yang akan diolah meliputi ekonomi,politik, kesehatan, olahraga dan kriminal saja.
3. Penelitian ini hanya mengolah data dalam bentuk bahasa Indonesia dan dari judul berita saja.

## **1.4 Tujuan dan Manfaat Penelitian**

### **1.4.1 Tujuan Penelitian**

Tujuan yang hendak dicapai dalam penelitian ini adalah sebagai berikut:

- a. Menggunakan *tools* rapid mineryang dapat mengimplementasikan metode Fuzzy C-means agar diperoleh *clusters* yang terbagi bagi menjadi beberapa tipe berita.

- b. Menganalisis *cluster* data dan menggunakan *cluster* tersebut untuk menentukan tipe.
- c. Mengetahui cara kerja dan tingkat akurasi metode Fuzzy C-means.

### **1.4.2 Manfaat Penelitian**

Adapun manfaat dalam penelitian ini adalah sebagai berikut:

- a. Hasil analisis *cluster* dapat menjadi tolak ukur dalam pemecahan masalah pengelompokan .
- b. Memberikan sumbangsih penelitian dalam bidang *clustering* khususnya pada objek permasalahan dokumen/artikel berita

## **1.5 Metodologi Penelitian**

### **1.5.1 Waktu Penelitian**

Dimulai dari Oktober 2019 sampai dengan Desember 2019.

### **1.5.2 Alat dan Bahan Penelitian**

Adapun alat dan bahan penelitian yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Perangkat Keras (*Hardware*)
  - a. *Laptop/Pc*
  - b. *Printer*
  - c. RAM 4 Gb
  - d. *Hardisk 500Gb*

- e. Spesifikasi Core i3
2. Perangkat Lunak (*Software*)
- a. Sistem Operasi *Windows 10*
  - b. Tools Rapid Miner 9.5
  - c. *Firefox Mozilla* digunakan untuk mencari literature dari *internet*.
  - d. *Microsoft Office* sebagai aplikasi pengolahan data untuk penulisan laporan.

### **1.5.3 Data Penelitian**

Adapun data penelitian ini yang dibutuhkan adalah 500 judul artikel berita yang disimpan dalam format txt yang mana sudah diunduh terlebih dahulu sebelumnya dari situs detik.com, kompas.com dan cnnindonesia.com beserta kategori ekonomi, kesehatan, kriminal, olahraga, dan politik

### **1.5.4 Metode Pengumpulan Data**

Pengumpulan data merupakan tahap penting dalam penulisan karya ilmiah. Pengumpulan data menurut Sugiyono (2012:137) dapat dilakukan dengan berbagai *setting*, berbagai sumber, dan berbagai cara dalam upaya mengumpulkan data. Adapun teknik atau metode pengumpulan data yang penulis lakukan metode observasi, yang mana merupakan cara mengumpulkan data dengan cara mengamati objek penelitian secara langsung. Peneliti mengadakan observasi langsung ke situs berita yang sudah ditentukan sebelumnya.

## **1.6 Sistematika Penulisan**

Sistematika penulisan skripsi ini memberikan penjelasan garis besar penelitian ini secara jelas supaya dapat lebih terlihat berhubungan yang disusun dalam kerangka bab dan sub-bab. Adapun sistematika penulisan dijabarkan dibawah ini sebagai berikut:

### **BAB I PENDAHULUAN**

Pada bab ini membahas tentang latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat, metode penelitian, metode pengumpulan data, dan sistematika penulisan.

### **BAB II TINJAUAN PUSTAKA**

Bab ini membahas tentang tinjauan pustaka yang digunakan dalam penelitian antara lain tinjauan umum meliputi definisi berita, *text mining* , langkah-langkah dalam *data mining* , metode Fuzzy C-Means dan perangkat lunak apa saja yang dipakai.

### **BAB III ANALISIS DAN PERANCANGAN**

Pada bab ini membahas tahapan-tahapan dari pra pengolahan dalam *data mining* dan korelasinya terhadap data yang berhasil dikumpulkan .

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini berisi hasil dan pembahasan tentang hasil data mining oleh aplikasi rapid miner , menampilkan hasil dan menganalisa hasilnya. Serta menceritakan kaitannya dengan clusternya..

## **BAB V KESIMPULAN DAN SARAN**

Bab ini menguraikan tentang kesimpulan dari keseluruhan bab-bab dan saran-saran dalam implementasi lebih lanjut.