

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Berdasarkan data dari Kementerian Komunikasi dan Informasi Indonesia yang diperoleh dari Lembaga Riset Pasar *E-Marketer*, populasi pengguna internet tanah air pada tahun 2017 mencapai 112 juta orang. Hal ini menunjukkan besarnya kebutuhan internet bagi penduduk Indonesia sebagai media untuk mendapatkan informasi.

Seiring dengan itu, berbagai bentuk dan jenis informasi tersedia melimpah di internet. Namun, setiap pengguna internet memiliki kebutuhan informasi tersendiri. Sebagian besar para pengguna hanya ingin memperoleh informasi dalam bidang tertentu saja. Membaca koleksi dokumen yang berjumlah banyak tentu akan membutuhkan waktu yang sangat lama.

Disamping itu, latar belakang penulis membuat Tugas Akhir ini adalah perkembangan *website* yang meningkat drastis sehingga mengakibatkan semakin pesat pula perkembangan artikel berita. Banyaknya informasi tersebut menyebabkan pengguna internet dapat mengalami kesulitan untuk menemukan informasi atau berita yang diinginkan, sehingga informasi tersebut tidak dapat dimanfaatkan secara optimal. Oleh karena itu dibutuhkan pengelompokan informasi agar lebih terstruktur sehingga informasi tersebut dapat dimanfaatkan secara maksimal.

Tahapan pertama yang dilakukan dalam mengolah informasi adalah *preprocessing*. *Preprocessing* adalah tahapan untuk mentransformasi data ke suatu format yang lebih mudah dan efektif untuk diproses oleh pemakai. Langkah-langkah dalam *preprocessing* terdiri dari *sentence*, *tokenization*, *stopword removal*, *stemming*, *TF (Term Frequency)*, *IDF (Inverse Document Frequency)*,  $TF*IDF$ .

Langkah selanjutnya adalah pengelompokan dokumen informasi dalam *data mining* dengan metode *clustering*. *Clustering* merupakan proses pengelompokan data atau objek ke dalam bentuk *cluster* (kelompok) sehingga setiap data dalam kelompok memiliki kemiripan variabel-variabel yang diteliti, dan didapatkan kemiripan objek dalam kelompok yang sama dibandingkan sesama objek dari kelompok yang berbeda. Tidak melakukan klasifikasi, estimasi, atau memprediksi nilai variabel target. Dan teknik ini hanya melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan. Terdapat banyak metode *clustering* yang dapat digunakan

untuk melakukan pengelompokan dokumen, diantaranya *partitional clustering* dan *hierarchical clustering*.

Masing-masing metode ini memiliki kelebihan dan kekurangan. *Partitional clustering* merupakan metode *clustering* yang prosesnya membagi objek menjadi beberapa partisi dan setiap objeknya hanya menjadi bagian dari sebuah *cluster*, sehingga tidak terjadi *overlapping*.

Sementara itu, *hierarchical clustering* merupakan metode *clustering* yang direpresentasikan dengan sebuah pohon biner yang disebut dendrogram pada metode *hierarchical clustering* objek *cluster*-nya dapat terjadi *overlapping*. Metode *hierarchical clustering* ini mengelompokkan data dengan cara membuat suatu hirarki berupa dendrogram dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak mirip di tempatkan pada hirarki yang berjauhan.

Penelitian ini menggunakan metode *K-Medoids*. Metode *K-Medoids* merupakan bagian dari *partitioning clustering*. Metode *K-Medoids* cukup efisien untuk *dataset* yang kecil. Langkah awal *K-Medoids* adalah mencari titik yang paling representatif (*medoids*) dalam sebuah *dataset* dengan menghitung jarak dalam kelompok dari semua kemungkinan kombinasi dari *medoids* sehingga jarak antar titik dalam suatu *cluster* kecil sedangkan jarak titik antar *cluster* besar.

Penelitian ini penulis melakukan penelitian terhadap data berita berbahasa inggris yang diambil dari beberapa situs *online*, kemudian disalin ke *notepad* dalam format *.txt*. Data berita tersebut berupa dokumen teks yang memiliki isi tulisan atau kumpulan kata berbahasa inggris. Metode pembobotan yang digunakan dalam penelitian ini ada metode TF-IDF. Metode tersebut digunakan untuk mendapatkan nilai *ranking* pada setiap kalimat. Sehingga dengan nilai *ranking* tersebut didapatkan kalimat utama dalam dokumen tersebut yang selanjutnya diproses untuk menghasilkan kesimpulan (*summary*) dari artikel/berita.

## 1.2 Perumusan Masalah

Berdasarkan permasalahan yang terdapat pada latar belakang, maka rumusan masalah yang terkait dalam penelitian ini adalah :

- a. Bagaimana menentukan kalimat utama dalam suatu dokumen berita menggunakan pembobotan *TF-IDF*.
- b. Bagaimana melakukan pengelompokkan dokumen berita menggunakan metode *K-Medoids*.

## 1.3 Tujuan

Tujuan yang dari penelitian ini adalah sebagai berikut:

- a. Menentukan kalimat utama dalam sebuah dokumen berita menggunakan pembobotan *TF-IDF*.
- b. Mengelompokkan dokumen berita dengan metode *K-Medoids*.

## 1.4 Batasan Masalah

Pada penelitian ini terdapat batasan masalah sebagai berikut :

- a. Menggunakan bahasa pemrograman *Java* dan *IDE Netbeans* sebagai *compiler*.
- b. *Dataset* yang digunakan adalah *dataset* berita *online* yang sudah disalin ke dalam *file notepad* dengan format *.txt*.
- c. Data teks menggunakan berita berbahasa inggris.
- d. Tidak menangani kesalahan pada penulisan kata dalam dokumen.
- e. Kualitas *cluster* tidak dibandingkan dengan algoritma lainnya.

## 1.5 Metodologi Penelitian

Untuk mendapatkan hasil tugas akhir yang sesuai dengan tujuan, terdapat metodologi penelitian sebagai berikut :

### a. Studi Literatur

Dilakukan untuk memahami tentang *k-medoids clustering* dan peringkasan dokumen otomatis dengan metode *TF-IDF* melalui buku, jurnal, paper, dan artikel.

### b. Analisa Perancangan

Merancang sistem *k-medoids clustering* dan peringkasan dokumen otomatis dengan metode *TF-IDF*, sehingga dapat menghasilkan keluaran yang diharapkan.

### c. Implementasi

Melakukan implementasi terhadap rancangan yang sudah dibuat.

### d. Pengujian dan Analisis

Melakukan pengujian dan analisis evaluasi terhadap sistem yang telah dibangun, supaya dapat diketahui kemampuan sistem yang sudah dibangun.

### e. Kesimpulan

Membuat kesimpulan dari hasil penelitian yang sudah dilakukan berdasarkan dengan tujuan yang sudah ditentukan sebelumnya.

### f. Penyusunan Laporan

Membuat dokumentasi penelitian dengan menyusun laporan penelitian.

## 1.6 Sistematika Penulisan

Penulisan tugas akhir ini dibagi dalam beberapa bagian. Setiap bagian menjelaskan langkah demi langkah dalam pengerjaan tugas akhir ini. Berikut adalah bagian tersebut:

### BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang penelitian serta tugas akhir, rumusan masalah, tujuan tugas akhir, dan batasan masalah dari judul tugas akhir. Serta metodologi penelitian dan sistematika penulisan yang digunakan dalam tugas akhir ini.

### BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang penjelasan sumber-sumber terkait yang digunakan dalam sistem yang dibuat, bersumber dari jurnal, buku, maupun artikel resmi dari internet.

### BAB III PERANCANGAN SISTEM

Bab ini membahas mengenai semua hal yang berkaitan dengan proses pemodelan dan perancangan sistem yang dibangun

### BAB IV PENGUJIAN DAN ANALISIS

Bab ini membahas pengujian-pengujian yang dilakukan pada sistem, baik dari *summarize* maupun *k-medoids clustering*. Dari hasil pengujian akan dilakukan analisis dan menarik kesimpulan dari hasil analisis tersebut. Serta akan membandingkannya dengan pendapat beberapa ahli.

### BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan akhir dari perancangan, analisis, dan pengujian yang diperoleh serta saran dan harapan untuk pengembangan penelitian lebih lanjut.