

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dengan perkembangan teknologi informasi dan komunikasi saat ini, berita dalam bentuk digital dapat ditemukan dengan mudah dan berjumlah sangat banyak yang berasal dari beberapa sumber. Banyaknya berita ini menyebabkan diperlukan suatu teknik pengolahan berita tersebut kedalam satu topik yang telah di definisikan sebelumnya agar dapat mempermudah pembaca memilih berita sesuai seperti topik yang di inginkan. Berdasarkan permasalahan tersebut upaya yang dapat dilakukan adalah menerapkan kategorisasi teks untuk mengelompokan berita berdasarkan kategorinya.

Kategorisasi pada berita memerlukan data yang terstruktur, namun data yang tersedia adalah data yang tidak terstruktur atau data yang masih mentah. Permasalahan ini sering di jumpai karena berita yang di dapatkan berasal dari internet. Oleh sebab itu pada kategorisasi teks dibutuhkan suatu proses yang merubah data tersebut menjadi data yang terstruktur dengan menggunakan metode TF-IDF.

Term frequency (TF) adalah pengukuran yang paling sederhana dalam metode pembobotan. Pada metode ini, masing-masing *term* diasumsikan mempunyai proporsi kepentingan sesuai jumlah terjadinya (munculnya) dalam teks (dokumen). *Term frequency* dapat memperbaiki nilai *recall* pada *information retrieval*, tetapi tidak selalu memperbaiki nilai *precision* (Tokunaga, Iwayama, 1994). Hal ini disebabkan *term* yang *frequent* cenderung muncul di banyak teks, sehingga term-term tersebut memiliki kekuatan pembeda (keunikan) yang kecil.

Inverse document frequency (IDF) adalah metode pembobotan term yang lebih condong (fokus) untuk memperhatikan kemunculan term pada keseluruhan kumpulan teks. Pada *IDF*, term yang jarang muncul pada keseluruhan koleksi teks dinilai lebih berharga. Nilai kepentingan tiap term diasumsikan berbanding terbalik dengan jumlah teks yang mengandung term tersebut (Tokunaga, Iwayama, 1994).

Kelebihan dari metode ini yaitu mengekstraksi kalimat-kalimat beranking tinggi untuk dimasukkan ke dalam rangkuman. Teks yang ada didalam sebuah berita dicari pembobotan menggunakan TF IDF sehingga akan ditemukan kata yang sering keluar untuk dijadikan kunci penentuan teks berita tersebut masuk kedalam kategori berita apa. Berdasarkan latar belakang diatas, maka tugas akhir ini mengambil topik tentang *automatic text summarization* dengan Judul “Pencarian Bobot Sebuah Teks Untuk Pengkategorian Berita Menggunakan TF-IDF”.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang masalah yang dijelaskan, maka rumusan masalah sebagai berikut :

Bagaimana mengkategorikan teks berita online menggunakan metode TF-IDF (*Term Frequency-Inversed Document Frequency*)?

1.3. Batasan Masalah

Adapun batasan masalah dalam pembuatan perangkat lunak ini adalah sebagai berikut :

1. Dokumen yang digunakan yaitu berita online yang bersumber dari media internet (*liputan6.com*).
2. Metode yang digunakan untuk menghitung bobot kalimat menggunakan metode TF-IDF (*Term Frequency-Inversed Document Frequency*).
3. Berita yang di uji memiliki batasan waktu pengujian yaitu mulai tanggal 11 agustus sampai 21 agustus 2018.

1.4. Tujuan Penelitian

Dari rumusan masalah diatas maka tujuan dari penelitian adalah sebagai berikut:

1. Membuat sebuah sistem pengkategorian berita online untuk menentukan kategori sebuah berita.
2. Mengimplementasikan metode TF-IDF (*Term Frequency-Inversed Document Frequency*) untuk melakukan proses pembobotan teks berita.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Dapat menentukan pengkategorian sebuah berita online
2. Memahami proses pembobotan sebuah teks berita online menggunakan metode tf-idf menggunakan aplikasi