

BAB 1

PENDAHULAN

1.1 Latar Belakang

Berita adalah informasi yang sering diakses oleh masyarakat melalui banyak media seperti media digital, surat kabar, televisi, dan media sosial [1]. Kebutuhan akan informasi teks berita dibutuhkan karena banyak orang saat ini ingin mengikuti perkembangan informasi di dunia edukasi, berita terbaru, keuangan, dan olahraga [2]. Pesatnya perkembangan teknologi internet saat ini membuat banyak media berita berkembang cukup pesat [3]. Perusahaan media massa surat kabar telah memanfaatkan teknologi internet untuk menyebarkan berita terkini secara online melalui media massa online [4][5].

Ratusan ribu berita ditulis dan dimuat setiap hari di portal berita Indonesia berbasis online, sehingga menyulitkan pembaca untuk menemukan topik berita yang ingin mereka baca [6]. dalam memudahkan pembaca untuk mendapatkan berita yang dicari, maka berita perlu diklasifikasikan menurut kategori masing – masing seperti pendidikan, berita terkini, keuangan, dan olahraga [7].

Sebelum mengklasifikasikan kategori berita, diperlukan suatu cara untuk mengolah berita tersebut. Salah satu cara untuk mengklasifikasikan kategori berita adalah dengan menggunakan Text Mining [8]. Penambangan teks adalah teknik klasifikasi penambangan data untuk memproses teks menggunakan komputer dengan tujuan menghasilkan analisis teks yang bermanfaat [9]. Pada tahapan Text mining memiliki tahapan yang sering disebut dengan praproses (*preprocessing*) data, dimana praproses (*preprocessing*) memiliki 5 tahapan yaitu *tokenizing*, *case folding*, *stopwords*, *stemming*, dan *TF-IDF Vectorizer* [10].

Penelitian tentang klasifikasi berita sudah pernah dilakukan oleh [11] dengan judul “Klasifikasi Online Dokumen Berita Dengan Menggunakan *Suffix Tree Clustering*” dengan melakukan pengujian terhadap waktu *generate suffix tree* dan respon *time* yang dibutuhkan aplikasi untuk memproses pencarian data berita menggunakan algoritma *clustering* mendapat nilai akurasi sebesar 80%.

Pada penelitian kedua yang dilakukan oleh [12] dengan judul ” Klasifikasi Berita Olahraga Menggunakan Metode *Naïve Bayes* dengan *Enhanced Confix Stripping Stemmer*” melakukan pengujian terhadap data latih dan data uji. Peneliti mengambil 151 berita latih dari 6 kategori yaitu, sepak bola, bola basket, raket, MotoGP, Formula 1 dan berita olahraga lainnya sedangkan data uji menggunakan dokumen tes berita olahraga diambil dari situs sport.detik.com secara acak dengan nomor 30 dokumen berita. Dari hasil uji coba yang dilakukan diperoleh hasil nilai akurasi sebesar 77% dengan error rate sebesar 23%.

Oleh karena itu perlu diperlukan pengembangan terhadap metode klasifikasi untuk mendapatkan nilai akurasi yang lebih baik lagi, maka peneliti tertarik melakukan penelitian yang berjudul “Analisis Komparasi Klafikasi Text Mining Berita Online Indonesia Menggunakan Algoritma *K- Nearest Neighbour* Dan *Random Forest*”, diharapkan penelitian yang dilakukan mendapatkan nilai akurasi diatas 80%.

1.2 Rumusan Masalah

Untuk membantu peneliti dalam melakukan penelitian maka ditentukan rumusan masalah, adapun rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana proses klasifikasi *Text Mining* berita menggunakan algoritma *K- Nearest Neighbour* dan *Random Forest*.
2. Berapa nilai akurasi yang diperoleh dalam melakukan klafisikasi *Text Mining* berita menggunakan algoritma *K-Nearest Neighbour* dan *Random Forest*.

1.3 Batasan Masalah

Untuk membuat penelitian ini tidak keluar dari topik akan diteliti maka peneliti membuat batasan masalah, adapun batasan masalah yang dibuat adalah sebagai berikut:

1. Berita yang diklasifikasikan adalah berita yang ada di negara indonesia dengan 5 kategori klasifikasi edukasi(*Education*), keuangan(*Finance*), terbaru(*hot*), terbaru(*news*), dan olahraga(*sport*).
2. Algoritma yang digunakan adalah *K-Nearest Neighbour* dan *Random Forest*.

1.4 Tujuan Penelitian

Penelitian dilakukan bertujuan untuk melakukan klasifikasi terhadap berita online sesuai dengan isi berita yang dikategorikan kedalam 5 kategori edukasi(*Education*), keuangan(*Finance*), terbaru(*hot*), terbaru(*news*), dan olahraga(*sport*).

1.5 Manfaat Penelitian

Diharapkan dengan adanya klasifikasi berita online penelitian ini para pembaca berita online dapat dengan mudah menemukan topik berita yang ingin dibaca.