

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan semakin berkembangnya teknologi, penyebaran informasi di internet berkembang dengan semakin pesat dan terus mengalami peningkatan. Salah satu bentuk informasi yang jumlahnya terus bertambah adalah berita (Wibisono, 2005)

Berita merupakan laporan tercepat mengenai fakta atau opini yang penting dan menarik minat bagi sejumlah besar orang. Aliran berita dalam jumlah besar tentu menjadi sumber informasi yang sangat berharga dan dapat dimanfaatkan untuk berbagai kepentingan. Salah satu pemanfaatannya adalah untuk melakukan monitoring isu.

Monitoring isu merupakan kegiatan untuk mengamati isu-isu terhangat yang sedang terjadi di masyarakat. Menurut Siti Meiningsih (Kominfo, 2014), dengan dilakukannya monitoring isu yang diambil dari berita di berbagai media, maka dapat dipetakan kebijakan apa yang akan diambil dalam tatanan pimpinan. Isu-isu yang terjadi di masyarakat sangat beragam, sehingga sangat tepat apabila isu-isu tersebut dikelola dan dianalisis untuk kemudian menjadi pijakan dalam membuat keputusan.

Analisis berita yang umumnya berbentuk teks dapat dilakukan dengan *text mining*. *Text mining* dapat didefinisikan sebagai suatu proses untuk mengekstrak informasi yang berguna dari suatu sumber data melalui identifikasi dan eksplorasi pola tertentu (Feldman & Sanger, 2006). Terdapat beberapa metode dalam *text mining*, salah satunya adalah klasifikasi.

Klasifikasi merupakan suatu proses pengelompokan dan pengkategorian suatu dokumen berdasarkan model terlatih yang sudah memiliki label sebelumnya (Miner, et al., 2012). Klasifikasi bermanfaat untuk mengelompokan data yang jumlahnya sangat banyak yang sulit dilakukan apabila diproses secara manual.

Sebelum penelitian ini, telah dilakukan penelitian serupa untuk mengelompokkan berita yang berasal dari media *online* www.kompas.com dengan menggunakan metode *Naive Bayes Classifier* (Wibisono, 2005). Dari penelitian ini dapat disimpulkan bahwa metode *Naive Bayes Classifier* cukup baik untuk mengelompokan dokumen teks Bahasa Indonesia ditunjukan dengan akurasi yang

mencapai 90,23%. Sama halnya dengan penelitian yang dilakukan oleh (Hamzah, 2012) yang mengelompokkan dokumen teks berita dan abstrak akademis dengan menggunakan *Naive Bayes Classifier*. Akurasi untuk dokumen berita mencapai 91% sedangkan pada dokumen akademik mencapai 82%.

Salah satu model *event Naive Bayes Classifier* adalah model multinomial. Dalam penelitian (Sumpeno & Destuardi, 2009) yang mengelompokkan teks Bahasa Indonesia ke dalam beberapa kelas emosi, dapat disimpulkan bahwa metode multinomial lebih baik digunakan untuk klasifikasi teks Bahasa Indonesia ditujukan dengan pengukuran F-Measure yang mencapai 62,15%.

Masalah yang umum ditemukan dalam proses klasifikasi maupun *clustering* dokumen adalah tingginya dimensi data, sehingga perlu dilakukan proses seleksi fitur untuk memilih beberapa fitur yang dapat digunakan untuk mewakili dokumen (Langgeni, Baizal, & A.W., 2010).

Berdasarkan pada penelitian-penelitian tersebut, maka penelitian ini dilakukan untuk menganalisis berita dengan cara pengelompokan berdasarkan kelas yang telah ditentukan menggunakan *Naive Bayes* model *multinomial*. Selain itu, dilakukan juga proses seleksi fitur menggunakan metode *Document Frequency Thresholding* dan pembobotan dengan menggunakan *Term Frequency-Inverse Document Frequency (TFIDF)* serta gabungan dari keduanya untuk mendapatkan perbandingan hasil yang terbaik.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka dapat disimpulkan suatu rumusan masalah yaitu bagaimana melakukan pengelompokan terhadap berita-berita yang terdapat pada beberapa media online menggunakan algoritma *Multinomial Naive Bayes*.

1.3 Batasan Masalah

Untuk memberi batas ruang lingkup penelitian, maka diberikan batasan masalah sebagai berikut :

1. Data berita yang digunakan merupakan berita-berita dari media online yang diambil dari Lembaga Pengolahan dan Penyedia Informasi (PPI), Dirjen Informasi dan Komunikasi Publik (IKP), Kementerian Komunikasi dan Informatika pada bulan Februari 2016 – Mei 2016.
2. Berita hanya dikelompokkan ke dalam satu kategori yang meliputi isu-isu untuk kategori bidang Polhukam, Ekonomi dan Kesra.
3. Berita yang digunakan adalah berita dengan teks Bahasa Indonesia.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menganalisis isu-isu yang terdapat pada berita di beberapa media online dengan teks klasifikasi menggunakan *Multinomial Naive Bayes*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian tugas akhir ini adalah dapat menemukan metode untuk klasifikasi dengan akurasi yang tinggi sehingga membantu lembaga penyedia informasi dalam melakukan monitoring berita-berita yang tersebar di beberapa media online.

1.6 Sistematika Penulisan

Laporan penelitian tugas akhir ini dilakukan dengan sistematika penulisan sebagai berikut :

1. BAB I Pendahuluan, terdiri dari pembahasan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
2. BAB II Tinjauan Pustaka, terdiri dari pembahasan teori-teori yang digunakan dalam penelitian, seperti penjelasan mengenai *text mining*, *text preprocessing*, *Document Frequency Thresholding*, *Term Frequency-Inverse Document Frequency (TFIDF)* dan juga algoritma *Multinomial Naive Bayes* sebagai

algoritma pengelompokan yang digunakan. Selain itu, dijelaskan juga mengenai penelitian terkait dan rencana penelitian yang dilakukan.

3. BAB III Metodologi Penelitian, berisi mengenai penjelasan langkah-langkah yang dilakukan dalam penelitian.
4. BAB IV Pembahasan, menjelaskan hasil dan pembahasan penelitian yang telah dilakukan.
5. BAB V Penutup, terdiri dari pembahasan kesimpulan sebagai jawaban dari permasalahan yang telah dirumuskan dan juga saran yang dapat dilakukan untuk penelitian selanjutnya.

