

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Internet adalah jaringan komunikasi global yang terbuka dan menghubungkan jutaan bahkan miliaran orang di seluruh belahan dunia baik dengan menggunakan komputer, telepon pintar maupun perangkat lainnya. Kebutuhan manusia zaman sekarang terhadap internet terus mengalami peningkatan. Terbukti dengan meningkatnya jumlah orang yang mengakses internet. Di Indonesia sendiri peningkatan jumlah orang yang mengakses internet naik 13% dibanding tahun 2018. Dengan begitu jumlah akses internet di Indonesia mencapai angka 56% dari seluruh jumlah penduduk Indonesia atau 150 juta orang. Angka ini akan terus meningkat tiap tahunnya sesuai dengan perkembangan zaman.

Berita adalah informasi berdasarkan fakta atau laporan mengenai suatu kejadian yang sedang atau telah terjadi dan dipublikasikan melalui media cetak, siaran, internet maupun dari mulut ke mulut. Dulu banyak instansi yang menyalurkan berita ke masyarakat melalui media televisi, koran, majalah dan radio. Namun akibat dari penggunaan internet yang terus meningkat sekarang banyak yang berpindah ke media *online* seperti *website* ataupun aplikasi berita yang selalu update setiap saat dan dapat diakses kapan saja dan dimana saja. Dengan banyaknya berita yang diunggah diinternet maka diperlukan pengorganisasian berita dengan cara mengklasifikasikan berita berdasarkan bahasan dalam berita tersebut. Hal ini bertujuan untuk memudahkan pembaca untuk menemukan berita yang diinginkan. selama ini editor berita melakukan klasifikasi secara manual sebelum berita tersebut diunggah. Namun dengan banyaknya berita yang masuk maka proses pengklasifikasian berita akan membutuhkan waktu yang lama. Maka dari itu dibutuhkan alat untuk melakukan klasifikasi secara otomatis.

Text Mining adalah salah satu bidang khusus dari data mining. Tujuan dari *Text Mining* adalah untuk mendapatkan informasi yang berguna dari suatu sumber data seperti dokumen. *Text Mining* dapat digunakan sebagai metode untuk mendapatkan informasi dari suatu berita, lalu informasi tersebut dapat digunakan untuk

menganalisa dan mengklasifikasi berita tersebut. Dengan menggunakan *Text Mining* ditambah penggunaan algoritma *Machine Learning* kita dapat melakukan klasifikasi berita dalam jumlah besar dalam waktu singkat.

Sebelumnya telah ada beberapa penelitian dalam melakukan klasifikasi konten berita diantaranya, Klasifikasi Konten Berita Digital Bahasa Indonesia Menggunakan *Support Vector Machines* (SVM) Berbasis *Particle Swarm Optimization* (PSO) (Nurhadi, 2015). Dalam penelitian yang dilakukan oleh Acmad Nurhadi pengklasifikasian berita menggunakan Teknik *Support Vector Machines* dengan menambahkan Teknik *Particle Swarm Optimization* sebagai metode pemilihan fitur akan diterapkan agar mendapatkan akurasi yang lebih akurat. Kemudian ada Klasifikasi Berita *Online* Menggunakan Metode *Support Vector Machine* dan *K- Nearest Neighbor* (Asiyah, Fithriasari, 2016). Pada penelitian yang dilakukan oleh Aisyah dan Fithriasari, mereka menggunakan 2 metode yang berbeda yaitu *Support Vector Machine* (SVM) dan *K- Nearest Neighbor* (KNN) untuk dikomparasikan agar diketahui metode mana yang lebih akurat dalam mengklasifikasi berita. Dan hasilnya SVM lebih baik dari pada KNN nilai akurasi, recall, precision dan F-Measure sebesar 93.2%, 93.2%, 93.63% dan 93.14%. Penelitian selanjutnya yang berjudul “*News Article Text Classification in Indonesian Language*” (Wongso, Luwinda, Trisnajaya, Rusli, & Rudy, 2017) mengkomparasikan metode *Support Vector Machine* (SVM) dan *Naïve Bayes* dengan menggunakan metode *feature selection* TF-IDF dan TF-IDF+SVD. Dan hasilnya menunjukkan bahwa kombinasi algoritma *Multinomial Naïve Bayes* dan TF-IDF menghasilkan nilai *precision* dan *recall* tertinggi.

Dalam penelitian ini penulis akan mengkomparasikan metode *feature selection* TF-IDF dan *Chi Square* pada sistem klasifikasi berita *online* di Indonesia dengan menggunakan metode *Multinomial Naïve Bayes* untuk mengklasifikasi teks berita *online* di Indonesia sesuai dengan kategorinya secara otomatis serta menentukan *feature selection* mana yang menghasilkan akurasi lebih baik.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana membangun sistem klasifikasi berita otomatis dengan menggunakan metode *Multinomial Naïve Bayes*?
2. Bagaimana hasil komparasi dari metode *feature selection* TF-IDF dan *Chi Square* pada klasifikasi berita *online* di Indonesia dengan metode *Multinomial Naïve Bayes*?

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut :

- a. Membangun aplikasi sistem klasifikasi berita yang dapat mengklasifikasi berita *online* di Indonesia secara otomatis.
- b. Mengetahui hasil perbandingan akurasi dari implementasi metode *Multinomial Naïve Bayes* dengan menggunakan *feature selection* TF-IDF dan *Chi Square* untuk klasifikasi berita *online* di Indonesia.

1.4 Manfaat

Manfaat dari penelitian ini adalah sebagai berikut :

- a. Mempermudah proses klasifikasi berita *online* di Indonesia dalam jumlah banyak.
- b. Mengetahui metode *feature selection* mana yang menghasilkan akurasi yang lebih tinggi antara TF-IDF dan *Chi Square* dalam klasifikasi berita menggunakan metode *Multinomial Naïve Bayes*.