

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini media *online* semakin banyak digunakan dan semakin berkembang. Salah satu manfaat dari media *online* adalah untuk membagikan karya kepada orang-orang berupa cerita pendek, novel, puisi atau karya tulis lainnya. Banyak situs-situs yang menyediakan berbagai macam cerita, salah satunya adalah <https://www.wattpad.com/>. Dalam situs ini pengunjung bebas membaca karya-karya tulisan orang lain. Selain membaca, pengunjung juga dapat menuliskan karya-karya untuk dipublikasikan kepada orang lain.

Setiap pembaca pasti memiliki kesukaan *genre* bacaan yang berbeda misalnya roman, horor, humor, spiritual, misteri dan sebagainya. *Genre* bacaan dapat dilihat dari isi *blurb*. Menurut Cambridge Dictionary arti dari *blurb* adalah deskripsi singkat dari buku, film atau lainnya yang ditulis oleh pembuat karya dan ditujukan untuk menarik minat orang supaya mau membeli, membaca, atau menonton karya tersebut. Seorang pembuat karya harus menuliskan *genre* dari karya yang dibuat ke dalam *genre* yang tepat supaya sesuai antara *genre* dan isi bacaan (Oman, 2017). Untuk dapat menentukan sebuah bacaan masuk ke dalam *genre* tertentu bukanlah perkara yang mudah apalagi bacaan yang tidak terlalu berbeda secara jelas (Ariadi & Fithriasari, 2015). Dengan menggunakan aplikasi penelitian ini *genre* dapat ditentukan secara otomatis.

Klasifikasi merupakan salah satu dari metode yang terdapat pada *text mining*. *Text mining* adalah suatu proses untuk mengekstrak informasi yang berguna dari suatu sumber data melalui identifikasi dan eksplosari pola tertentu (Feldman & Sanger, 2007). Klasifikasi dapat dimanfaatkan untuk mengelompokkan data yang jumlahnya banyak. Ada beberapa jenis algoritma yang dapat digunakan untuk klasifikasi teks yaitu NBC (*Naïve Bayes Classifier*), SVM (*Support Vector Machine*), dan *decission tree* (Feldman & Sanger, 2007). Penelitian ini menggunakan metode klasifikasi *Naïve Bayes Classifier (NBC)* karena memiliki beberapa kelebihan yaitu sederhana, cepat, dan memiliki akurasi yang tinggi (Saptono, Sulisty, & Trihabsari, 2016). Salah satu model *event Naïve Bayes Classifier* adalah model multinomial. Dalam sebuah penelitian yang menggunakan model multinomial tentang pengelompokan teks Bahasa Indonesia ke dalam beberapa emosi menghasilkan pengukuran F-Measure mencapai 62,15% (Rahman, Wiranto, & Doewes, 2017). Penelitian yang dilakukan oleh Amelia Rahman, Wiranto dan Afrizal Doewes (2017) tentang klasifikasi berita *online* menggunakan *Multinomial Naïve Bayes* menghasilkan rata-rata akurasi sebesar 86,62%.

Dalam penelitian ini menggunakan *Multinomial Naïve Bayes Classifier* yaitu dengan memperhitungkan frekuensi setiap kata yang muncul pada dokumen. Tahapan awal adalah *text preprossesing* yang terdiri dari empat langkah yaitu *case folding, tokenizing, filtering, dan stemming*. Kemudian dilakukan pembobotan kata menggunakan *Term Frequency Inverse Document Frequency (TF IDF)* yang bertujuan untuk memberikan bobot pada setiap kata. Dan untuk memperoleh hasil klasifikasi yang optimal maka digunakan reduksi fitur *Document Frequency (DF)*

Thresholding. *DF Thresholding* adalah salah satu teknik reduksi fitur yang paling sederhana namun memiliki kinerja yang baik.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, terdapat permasalahan yang dirumuskan sebagai berikut:

- a. Bagaimana mengimplementasi *Multinomial Naïve Bayes Clasifier* dan reduksi fitur *DF Thresholding* untuk klasifikasi genre berdasarkan *blurb*?
- b. Bagaimana hasil performa dari metode *Multinomial Naïve Byes Clasifier* dan reduksi fitur menggunakan *DF Thresholding* untuk klasifikasi genre berdasarkan *blurb*?

1.3 Batasan Masalah

Batasan masalah dalam pembuatan sistem ini adalah sebagai berikut:

- a. Data yang digunakan dalam penelitian ini diperoleh dari aplikasi Wattpad atau dapat juga diakses di <https://www.wattpad.com/>.
- b. Klasifikasi dokumen akan dibagi menjadi lima *genre* yaitu roman, horor, humor, misteri dan spiritual.
- c. Metode klasifikasi yang digunakan adalah metode *Multinomial Naïve Bayes Clasifier*.
- d. Reduksi fitur menggunakan metode *DF Thresholding* dan pembobotan menggunakan teknik *TF IDF*.

- e. Aplikasi yang dibuat berbasis web dengan menggunakan bahasa pemrograman PHP.

1.4 Tujuan Penelitian

Tujuan dari tugas akhir ini adalah membangun suatu sistem untuk klasifikasi *genre* berdasarkan *blurb* cerita pada media berbagi cerita *online* dengan menggunakan metode *Multinomial Naïve Bayes Clasifier*.

1.5 Manfaat Penelitian

Penelitian ini dilakukan dengan harapan dapat memberikan manfaat kepada pembuat karya supaya dengan otomatis mengklasifikasi karya yang dibuat dengan memasukkan dalam sebuah sistem dengan menggunakan metode tertentu.