

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Industri film merupakan salah satu industri yang ada di Indonesia yang terus mengalami perkembangan. Pada tahun 2019, *Global Box Office* berhasil meraih penghasilan sebesar 42,2 miliar dolar yang merupakan penghasilan yang lebih besar ketimbang tahun-tahun sebelumnya (Rivkin, 2019). Untuk industri hiburan *home/mobile* yang meliputi jasa *streaming* memiliki perkembangan sebesar 14 persen secara global (Rivkin, 2019). Hal tersebut membuktikan bahwa sepanjang tahunnya, industri hiburan terutama film terus mengalami perkembangan.

Sebelum film ditonton, melihat *review* yang beredar terkait film tersebut menjadi salah satu kebiasaan dari masyarakat sebelum menonton film. Diharapkan dengan melihat *review*, membuat orang menjadi lebih tertarik untuk menonton film terkait. Namun, tidak sedikit di antara *review* tersebut mencakup hal-hal atau adegan penting terkait yang ada pada film atau yang dapat kita sebut dengan *spoiler*. Adanya *spoiler* menghilangkan minat seseorang untuk menikmati konten dan karena itu menjadikannya dampak buruk bagi beberapa orang (Richler, 2013) sehingga mempengaruhi kesan penonton terhadap film dan akhirnya berdampak pada nilai jual film.

Di era digital, penyebaran informasi bergerak begitu cepat sehingga penyebarannya dapat menyebar luas dalam waktu yang singkat sehingga perlu dilakukan pencegahan atau penanganan agar dampak dari penyebaran *spoiler* tersebut dapat berkurang. Beberapa penelitian telah dilakukan sebagai penanganan seperti *topic model based* pada *Latent Dirichlet Allocation* (LDA) menggunakan

Linguistic Dependencies untuk menghadapi fenomena *spoiler* pada komentar *online* (Guo & Ramakrishnan, 2010). Penggunaan *Neural Network Machine Learning Architecture* juga digunakan untuk mendeteksi plot yang sensitif pada dokumen *review film* (Nakashole, 2019).

Diluar dari penelitian yang telah disebutkan, penggunaan klasifikasi teks dapat dilakukan untuk mendeteksi teks yang mengandung *spoiler*. Klasifikasi teks adalah proses labelisasi atau kategorisasi untuk dokumen yang belum pernah diolah atau dilihat (Drame, Mougine, & Diallo, 2016). Klasifikasi teks dapat dilakukan dengan *machine learning* dan karenanya terdapat beberapa algoritma yang dapat diterapkan. Salah satu algoritma yang dapat digunakan yaitu teknik *bagging* dan juga *Multinomial Naïve Bayes*.

Teknik *bagging* dipilih berdasarkan pertimbangan bahwa teknik ini sangat membantu terutama mengatasi masalah ketidak seimbangan kelas klasifikasi dan telah berhasil meningkatkan akurasi untuk klasifikasi tertentu pada *dataset* buatan dan yang sebenarnya (Prasetio & Pratiwi, 2015). *Multinomial Naïve Bayes* sendiri merupakan salah satu varian dari *Naïve Bayes* dan sudah dipakai diberbagai penelitian sebagai algoritma untuk klasifikasi teks (Xu, Li, & Zhang, 2017). Untuk akurasi sendiri, penelitian lain sudah dilakukan untuk menguji *Multinomial Naïve Bayes*. *Multinomial Naïve Bayes* mampu mencapai akurasi 89,58% untuk sistem klasifikasi surat yang merupakan salah satu penelitian mengenai klasifikasi teks (Kalokasari, Shofi, & Setyaningrum, 2017). Berdasarkan pertimbangan tersebut, algoritma ini dinilai dapat digunakan untuk klasifikasi teks pada penelitian ini.

1.2 Rumusan Masalah

Berdasarkan masalah yang diuraikan pada latar belakang dan metode yang dipilih untuk penelitian ini, maka dapat disusun rumusan masalah sebagai berikut:

- 1) Bagaimana cara mengimplementasikan algoritma *Multinomial Naïve Bayes* dan Teknik *Bagging* untuk klasifikasi teks pada data *spoiler* film?
- 2) Bagaimana tingkat akurasi hasil test dari teknik *Bagging* yang dikombinasikan dengan *Multinomial Naïve Bayes* yang diukur dengan metrik *F1-Score* untuk klasifikasi teks pada data *spoiler* film?

1.3 Batasan Masalah

Penelitian ini sendiri memiliki beberapa batasan masalah yang terbagi menjadi beberapa bagian yaitu:

- 1) *Dataset* yang digunakan adalah *IMDB Spoiler Dataset* tersedia pada situs *web kaggle*. *Dataset* tersedia di <https://www.kaggle.com/rmisra/imdb-spoiler-dataset>.
- 2) *Dataset* yang digunakan tersedia dalam teks berbahasa *inggris*.
- 3) *Dataset* terbatas pada judul film pada tahun 1998 sampai 2018.
- 4) *Dataset* yang digunakan telah dilabeli dengan label *spoiler* dan *not spoiler*.
- 5) Pengukuran performa menggunakan *F1-Score*.
- 6) Metode ekstrasi fitur menggunakan metode *TF-IDF*.

1.4 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang ada, maka tujuan penelitian dapat diuraikan sebagai berikut:

- 1) Mengetahui cara mengimplementasikan algoritma *Multinomial Naïve Bayes* dengan teknik *bagging* dalam klasifikasi teks mengandung *spoiler* film.
- 2) Mengetahui pengaruh teknik *bagging* terhadap nilai evaluasi model *multinomial naïve bayes* yang diukur dengan metrik pengukuran *F1-Score* untuk klasifikasi teks mengandung *spoiler* film.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dengan tercapainya tujuan penelitian adalah sebagai berikut:

- 1) Mengetahui tingkat akurasi dari model yang dihasilkan berdasarkan algoritma *Multinomial Naïve Bayes* dan teknik *bagging*.
- 2) Sebagai referensi untuk bidang penelitian serupa ke depannya untuk menciptakan system yang lebih baik.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan laporan skripsi ini dijabarkan sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini menjelaskan latar belakang, rumusan masalah, batasan masalah, tujuan, dan manfaat dari penelitian ini.

BAB 2 LANDASAN TEORI

Bab ini menjelaskan beberapa teori yang digunakan dalam penelitian ini. Teori yang digunakan antara lain text classification, text preprocessing, bagging, multinomial naïve bayes, dan f1- score.

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah yang diterapkan dalam merancang dan membangun sistem penelitian. Susunan sistem dijelaskan dalam bentuk flowchart yang terbagi menjadi flowchart utama, flowchart pipeline model, flowchart preprocessing, flowchart tf-idf, flowchart bagging with multinomial naïve bayes.

BAB 4 HASIL DAN DISKUSI

Bab ini menjelaskan mengenai spesifikasi sistem yang digunakan untuk penelitian dan juga implementasi algoritma bagging dan multinomial naïve bayes untuk membentuk model machine learning untuk deteksi spoiler pada dataset review film.

BAB 5 SIMPULAN DAN SARAN

Bab ini berisi tentang hasil uji coba dan analisis yang telah didapatkan dari penelitian yang dilakukan sekaligus saran untuk mengembangkan sistem menjadi lebih baik.