

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan populasi pengguna internet semakin meningkat dengan cepat. Populasi tersebut mengakibatkan banyaknya opini di media. Opini orang lain di media sangat penting untuk diketahui polaritasnya karena opini dapat membantu dalam membuat suatu keputusan. Banyak situs yang menyediakan *review* terhadap suatu produk yang dapat mencerminkan opini pengguna, salah satunya adalah situs *Internet Movie Database* (IMDb). IMDb merupakan salah satu situs yang menyediakan database *review movie*. Opini terhadap *movie* sangat penting terhadap pembuatan keputusan.

Klasifikasi sentimen merupakan salah satu sistem yang mampu mendeteksi polaritas dari suatu opini *movie* dengan otomatis dengan bantuan komputerisasi. Namun yang menjadi tantangan terbesar dalam klasifikasi ini adalah pertama, dalam dataset terdapat banyak atribut. Atribut tersebut terdiri dari atribut penting dan atribut *noisy* pada dataset. Hal tersebut mempengaruhi performansi akurasi yang akan dihasilkan oleh suatu metode klasifikasi. Kedua, negasi dimana suatu kata bisa saja polaritasnya positif namun pada konteks teks yang berbeda polaritasnya bisa negatif, seperti kata “*bad*” pada kalimat dengan dua konteks teks yang berbeda yaitu “*the movie is bad*” dan “*the movie is not bad*”. Negasi tersebut merupakan suatu hal yang harus ditangani, karena negasi merupakan salah satu permasalahan yang paling umum yang dapat mengubah polaritas dari suatu teks. Ketiga, format teks yang tidak terstruktur. Ketiga masalah tersebut sangat mempengaruhi terhadap penentuan polaritas terhadap suatu opini.

Sebelumnya telah dilakukan penelitian oleh peneliti [12] pada *movie review* dengan menggunakan metode *mechine learning* yaitu multinomial naïve bayes,

maximum entropy, support vector machine, stochastic gradient descent. Setiap metode tersebut dikombinasikan dengan menggunakan n-gram. Hasil diperoleh bahwa dengan kombinasi n-gram dapat meningkatkan nilai akurasi terhadap setiap metode yang diimplementasikan. Oleh sebab itu penulis mencoba untuk menyelesaikan masalah pada data *movie review* dengan menggunakan metode kombinasi lain dengan menyelesaikan permasalahan diatas untuk meningkatkan nilai akurasi terhadap metode multinomial naïve bayes. Adapun kombinasi yang dilakukan adalah dengan menerapkan proses *preprocessing*, *negation handling* dan melakukan pembobotan pada setiap fitur. Proses klasifikasi sentimen dengan metode multinomial naïve bayes akan dibagi menjadi dua kelas yaitu kelas sentimen positif dan kelas sentimen negatif. *Preprocessing* dilakukan sebelum dilakukan tahap klasifikasi. Tujuan dari *preprocessing* ini adalah untuk mengurangi atribut *noisy*. Dalam memecahkan permasalahan negasi maka lingkup negasi yang akan diselesaikan yaitu penggunaan kata “not”, “no”, “n’t”. Untuk mengatasi negasi tersebut penulis akan melakukan dua teknik dalam memecahkan permasalahan negasi. Teknik yang dilakukan adalah pertama, penanganan negasi berdasarkan *punctuation* dimana ketika ditemukan kata “not”, “no”, “n’t” maka akan digabungkan dengan setiap kata setelah negasi sampai ditemukan *punctuation*. Teknik kedua, yaitu dengan menggunakan teknik yang sudah dilakukan oleh peneliti [8] yaitu dengan metode *rules* dengan bantuan identifikasi kata dengan *pos tagging*. TF-IDF digunakan untuk melakukan pembobotan pada setiap fitur. Diharapkan dengan menggunakan teknik-teknik tersebut dapat meningkatkan hasil nilai akurasi dengan metode klasifikasi yang digunakan yaitu multinomial naïve bayes.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas dalam penelitian ini adalah :

- a. Bagaimana mengekstrak fitur untuk mengurangi atribut *noisy* dengan proses *preprocessing* dan apa pengaruh fitur yang di ekstrak tersebut terhadap klasifikasi *movie review*?

- b. Bagaimana merancang dan menganalisis pengaruh negasi *syntactic* yaitu mencakup negasi dengan fitur “*not*”, “*no*”, “*n't*” ?
- c. Bagaimana pengaruh pemberian bobot terhadap fitur yang digunakan pada proses klasifikasi?

1.3 Tujuan Penelitian

Tujuan pengerjaan tugas akhir ini adalah, sebagai berikut:

- a. Merancang dan menganalisis pengaruh ekstraksi fitur yang dilakukan dengan teknik *preprocessing* terhadap klasifikasi *movie review*.
- b. Merancang dan menganalisis teknik *negation handling* dengan lingkup negasi “*not*”, “*no*”, “*n't*” pada *movie review*.
- c. Menganalisis pengaruh bobot dengan TF-IDF pada setiap fitur dalam proses klasifikasi pada *movie review*.

1.4 Batasan Masalah

Dalam pengerjaan tugas akhir ini, ada beberapa batasan masalah dalam menyelesaikan masalahnya yaitu :

- a. Data yang digunakan adalah *movie review* dengan text bahasa *English*.
- b. Fokus utama dalam penelitian ini adalah menganalisis bagaimana pengaruh performansi yang dihasilkan oleh metode *preprocessing*, *negation handling* dan TF-IDF dengan metode klasifikasi multinomial naïve bayes.
- c. Klasifikasi opini terdiri dari dua kelas polaritas yaitu kelas positif dan negatif.

1.5 Metodologi Penelitian

Tahapan metode penelitian yang dilakukan dalam menyelesaikan permasalahan ini adalah sebagai berikut :

1. Tahap identifikasi masalah

Pada tahap ini, melakukan identifikasi masalah terhadap judul Tugas Akhir (TA) yang akan diangkat. Suatu permasalahan diidentifikasi dengan sebaik mungkin supaya solusi yang akan diusulkan untuk memecahkan masalah dapat diselesaikan sesuai dengan harapan yang di inginkan.

2. Tahap studi literature

Pada tahap ini, penulis mencari dan mengumpulkan sumber referensi yang berkaitan dengan pengerjaan TA yaitu mencari permasalahan dan solusi yang akan digunakan. Dalam mencari referensi penulis melakukan pencarian dari internet, jurnal, buku dan hasil diskusi dengan dosen pembimbing. Tujuan dari tahap ini adalah untuk menemukan solusi yang tepat dalam menyelesaikan masalah yang telah teridentifikasi.

3. Tahap pengambilan data

Pada tahap ini ditentukan data yang akan digunakan dalam pengerjaan klasifikasi sentimen ini.

4. Tahap perancangan sistem

Pada tahap ini, penulis melakukan proses-proses perancangan berdasarkan metode solusi yang dipilih dalam menyelesaikan setiap permasalahan yang dirumuskan. Perancangan yang akan dilakukan adalah merancang alur-alur dalam menyelesaikan setiap permasalahan berdasarkan metode solusi yang digunakan untuk mencapai tujuan dalam pengerjaan masalah ini.

5. Tahap implementasi sistem

Pada tahap ini, penulis mengimplementasikan perancangan yang telah dibuat pada tahap sebelumnya.

6. Tahap pengujian dan evaluasi sistem

Pada tahap ini, penulis melakukan pengujian terhadap implementasi yang telah dilakukan pada tahap implementasi sebelumnya. Tujuan dari pengujian ini adalah untuk mengetahui sejauh mana tingkat akurasi yang diperoleh ketika menggunakan metode yang dipilih.

7. Tahap analisis pengujian

Pada tahap ini, penulis akan melakukan analisis dan penarikan kesimpulan terhadap metode yang digunakan penulis berdasarkan hasil pengujian.

8. Tahap penulisan laporan

Pada tahap ini, penulis akan menulis laporan dan pengumpulan dokumen yang dibutuhkan berdasarkan hasil penelitian yang telah dilakukan.

1.6 Sistematika Penulisan

Adapun sistematika penulisan dalam pengerjaan Tugas Akhir ini adalah sebagai berikut:

a. BAB I

Bab ini berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metode penelitian dan sistematika penulisan.

b. BAB II

Bab ini berisi penjelasan singkat mengenai tinjauan pustaka dan penelitian terkait yang mendukung dalam penyelesaian masalah yang dibahas dalam pengerjaan Tugas Akhir.

c. BAB III

Bab ini berisi sistematika perancangan sistem yang akan dibangun dalam membuat klasifikasi sentiment pada movie review.

d. BAB IV

Bab ini berisi hasil pengujian sistem berdasarkan rancangan yang telah dibuat serta analisis dari pengujian.

e. BAB V

Bab ini berisi kesimpulan yang didapat dari hasil analisis sistem dan saran untuk pengembangan selanjutnya.