

# BAB 1 PENDAHULUAN

Bab ini akan menjelaskan latar belakang dari penelitian yang memberikan gambaran awal mengenai analisis sentimen dan perannya dalam perkembangan teknologi informasi. Tujuan dan ruang lingkup pengerjaan tugas akhir memberikan penjelasan mengenai hasil yang ingin diketahui serta batasan dalam pengerjaan. Tahapan dalam metodologi penelitian dan sistematika penulisan laporan akan dibahas pada dua subbab terakhir.

## 1.1 Latar Belakang

Kebutuhan akan informasi mendorong perkembangan penelitian dan teknologi yang dapat menjawab kebutuhan tersebut. Informasi yang dibutuhkan mengalami perkembangan dari informasi yang bersifat umum ke informasi yang bersifat lebih khusus dan spesifik. Perolehan informasi yang tepat dan cepat akan membantu dalam kemajuan dari suatu organisasi untuk dapat melakukan perubahan yang menjawab permasalahan yang dihadapi. Penelitian dalam analisis sentimen didorong oleh suatu pemikiran bahwa informasi berupa sentimen dari suatu data merupakan hal yang penting dan dibutuhkan. Sentimen berhubungan dengan penilaian terhadap suatu konteks atau wacana. Sentimen positif menyatakan pemberian nilai yang baik pada konteks dalam teks dan sentimen negatif menyatakan kebalikannya.

Sebagai contoh, analisis sentimen dapat diaplikasikan pada perusahaan yang mengeluarkan suatu produk dan menyediakan layanan untuk menerima pendapat (*feedback*) dari konsumen untuk produk tersebut. Analisis sentimen diaplikasikan untuk mengelompokkan *feedback* positif dan negatif dari konsumen sehingga mempercepat dan mempermudah tugas perusahaan untuk meninjau kembali kekurangan produk mereka, berdasarkan *feedback* negatif yang diterima dari konsumen, atau untuk mengetahui tingkat penerimaan konsumen terhadap produk tersebut.

Kebutuhan informasi yang khusus tersebut yang mendorong penelitian analisis sentimen yang ada, sekaligus yang memotivasi penelitian dalam tugas akhir ini untuk melakukan analisis sentimen pada domain bahasa Indonesia. Penelitian analisis sentimen bahasa Indonesia diharapkan dapat membantu dalam menjawab kebutuhan informasi sentimen yang memperkaya keberagaman dan ketersediaan informasi untuk bahasa Indonesia.

## **1.2 Permasalahan**

Permasalahan dalam tugas akhir ini adalah belum diketahuinya bagaimanakah penerapan analisis sentimen untuk bahasa Indonesia. Penelitian analisis sentimen dalam bahasa Inggris telah banyak dilakukan dan beberapa metode juga telah dikembangkan. Sementara untuk bahasa Indonesia, belum diketahui apakah analisis sentimen dapat dilakukan menggunakan metode-metode tersebut atau apakah diperlukan pencarian metode tertentu yang lebih tepat diterapkan pada domain bahasa Indonesia. Permasalahan lain yang dihadapi adalah keterbatasan sumber daya berupa data dalam bahasa Indonesia, yang diperlukan untuk melakukan analisis sentimen. Pada tugas akhir ini, dokumen berbahasa Indonesia didapat dari hasil penerjemahan otomatis dokumen berbahasa Inggris seperti yang akan dibahas di subbab 4.1.2.

## **1.3 Tujuan**

Tujuan utama dari tugas akhir ini adalah untuk mengetahui kinerja dari penerapan metode *machine learning* Naive Bayes, Maximum Entropy, dan Support Vector Machine dalam melakukan analisis sentimen untuk dokumen berbahasa Indonesia hasil penerjemahan otomatis dibandingkan dengan dokumen berbahasa Inggris, dengan mencoba mengaplikasikan kembali percobaan yang dilakukan dalam (Pang, Lee, & Vaithyanathan, 2002). Selain itu, tugas akhir ini juga bertujuan untuk mengetahui kinerja analisis sentimen tiap metode dengan melihat perbandingan akurasi analisis sentimen dari aspek bahasa (Inggris dan Indonesia), fitur *machine learning* yang digunakan, dan nilai atau informasi fitur yang dipakai.

## 1.4 Ruang Lingkup

Secara umum, dasar pengerjaan tugas akhir ini mengikuti percobaan yang dilakukan pada (Pang, Lee, & Vaithyanathan, 2002). Ruang lingkup pengerjaan dari tugas akhir ini adalah sebagai berikut:

1. Analisis sentimen pada data tekstual menggunakan dokumen *review* dalam domain film. Data *review* film didapat dari [http://www.cs.cornell.edu/people/pabo/movie-review-data/mix20\\_rand700\\_tokens.zip](http://www.cs.cornell.edu/people/pabo/movie-review-data/mix20_rand700_tokens.zip). Data yang digunakan mencakup data bahasa Inggris dan data bahasa Indonesia hasil penerjemahan otomatis, menggunakan kamus bilingual dan Transtool v4.1.
2. Fitur yang digunakan adalah fitur *unigram* dengan melakukan variasi pada nilai fitur dan pemilihan fitur. Fitur bigram dan kombinasi fitur dengan *part-of-speech* tidak digunakan.
3. Analisis sentimen dengan metode *machine learning* Naive Bayes, Maximum Entropy, dan Support Vector Machine menggunakan *tools* yang sudah tersedia (subbab 4. 4) dan mengklasifikasikan dokumen ke dalam dua kelas sentimen, positif dan negatif. Pekerjaan yang dilakukan mencakup penerjemahan otomatis data bahasa Inggris, pemilihan fitur, persiapan data masukan untuk masing-masing *tools*, dan pemanfaatan *tools* atau *library* yang tersedia untuk analisis sentimen.
4. Pembuatan *baseline* untuk analisis sentimen pada dokumen berbahasa Indonesia (subbab 3. 7 dan subbab 4. 5).

## 1.5 Metodologi Penelitian

Metodologi yang digunakan dalam pengerjaan tugas akhir ini adalah metode eksperimental dan dilaksanakan dalam tahapan-tahapan sebagai berikut:

1. Studi literatur - Pencarian informasi mengenai analisis sentimen serta penelitian yang telah dilakukan dalam bidang tersebut. Pada studi literatur juga akan dilakukan pembelajaran dari metode-metode *machine learning* yang digunakan untuk analisis sentimen.

2. Perancangan – Melakukan perancangan percobaan dengan mempersiapkan data, penentuan variabel dalam percobaan, persiapan *baseline*, dan perancangan proses analisis sentimen menggunakan *machine learning*.
3. Perumusan hipotesa – Hipotesa dirumuskan sebelum melakukan implementasi percobaan. Hipotesa yang digunakan adalah bahwa akurasi analisis sentimen pada dokumen berbahasa Indonesia akan memiliki nilai akurasi yang lebih besar dari *baseline* yang dirancang dan mendekati nilai akurasi analisis sentimen pada dokumen berbahasa Inggris, tetapi tidak lebih besar. Hipotesa ini dibuat dengan pemikiran bahwa dokumen hasil penerjemahan otomatis memiliki kualitas yang rendah dan bukan merupakan hasil penerjemahan yang sempurna dari dokumen asli berbahasa Inggris, sehingga tidak akan memberi hasil yang melebihi akurasi analisis sentimen pada data bahasa Inggris.
4. Implementasi - Pada bagian implementasi dilakukan penerapan dari perancangan yang dilakukan dengan melakukan pengolahan data, pembuatan *baseline*, dan penggunaan metode *machine learning* untuk analisis sentimen.
5. Analisis hasil – Melakukan perbandingan akurasi analisis sentimen dengan *machine learning* terhadap *baseline* serta perbandingan akurasi yang didapat dengan metode *machine learning* dilihat dari aspek bahasa, metode, fitur, dan nilai fitur yang digunakan.

## **1.6 Sistematika Penulisan**

Sistematika penulisan laporan mengikuti tahapan-tahapan yang dilakukan untuk menyelesaikan tugas akhir, sebagai berikut:

### – BAB 1 PENDAHULUAN

Memberikan penjelasan mengenai latar belakang, permasalahan, dan tujuan dari tugas akhir. Pada bagian pendahuluan juga diberikan ruang lingkup, metodologi, dan sistematika penulisan tugas akhir ini.

### – BAB 2 LANDASAN TEORI

Memberikan penjelasan secara singkat mengenai analisis sentimen dan penelitian yang telah dilakukan dalam analisis sentimen. Dasar teori dari metode *machine learning* Naive Bayes, Maximum Entropy, dan Support

Vector Machine diberikan untuk memberikan gambaran dan cara kerja masing-masing metode tersebut.

– BAB 3 PERANCANGAN

Menjelaskan perancangan percobaan berupa tahapan-tahapan untuk melakukan analisis sentimen pada dokumen.

– BAB 4 IMPLEMENTASI

Menjelaskan detail implementasi untuk melakukan analisis sentimen pada dokumen berbahasa Indonesia dan Inggris, mulai dari penerjemahan dokumen hingga proses analisis dengan *machine learning* dan pembuatan *baseline*. Pada bagian implementasi juga diberikan penjelasan dari variasi data dan fitur yang digunakan dalam percobaan.

– BAB 5 HASIL DAN PEMBAHASAN

Memberikan hasil dari percobaan analisis sentimen yang dilakukan. Hasil yang ditunjukkan berupa hasil dari *baseline* dan hasil analisis sentimen menggunakan *machine learning* yang dilihat dari beberapa aspek.

– BAB 6 PENUTUP

Berisi kesimpulan dan kendala dari analisis sentimen untuk dokumen berbahasa Indonesia serta saran untuk penelitian lanjutan dalam topik ini.