BAB 1
PENDAHULUAN

1.1 Latar Belakang

Analisis sentimen adalah salah satu area penelitian yang populer dalam *natural language processing*, *data mining*, *information retrieval*, dan *web mining* yang digunakan untuk menganalisis atau mengidentifikasi opini, sentimen, dan emosi seseorang dalam menyikapi suatu topik tertentu. Analisis sentimen mengklasifikasi polaritas menjadi 3 yaitu positif, negatif, dan netral [1]. Di dalam analisis sentimen setiap kalimat hanya dapat diidentifikasi satu polaritas, meskipun ada kemungkinan satu kalimat memiliki polaritas yang berbeda, untuk itu dibutuhkan analisis sentimen berdasarkan aspek. Analisis sentimen berdasarkan aspek bertujuan untuk mengidentifikasi polaritas pada aspek tertentu dengan mengklasifikasi tiap kata yang memiliki kategori yang sama. Setiap kata pada suatu kalimat memiliki aspek dan polaritas sentimen yang berbeda-beda dan membuat proses analisis sentimen lebih akurat [2].

Opini sangat penting bagi sebagian besar aktivitas manusia karena dapat mempengaruhi sentimen bisnis, politik, dan emosi yang berdampak pada sistem sosial dan politik. Dengan demikian perlu untuk mengambil dan mempelajari opini di media sosial. Namun, dikarenakan teks opini di media sosial memiliki jumlah yang besar, sulit untuk mengidentifikasi polaritas setiap opini. Oleh karena itu dibutuhkan analisis sentimen dengan pelabelan data secara otomatis untuk menemukan polaritas opini dan mengklasifikasikan opini tersebut sebagai positif atau negatif secara cepat. Selain itu data yang didapatkan bisa langsung ditampilkan dengan cepat dalam bentuk grafik dan tabel [3]. Pada penelitian [4] yang dilakukan oleh M. S. Adhi, M. Z. Naf'an, dan E. Usada, analisis sentimen dengan proses pelabelan kata dilakukan secara manual yang menghasilkan nilai akurasi sebesar 70% menggunakan algoritma *Naïve Bayes Classifier* memiliki kekurangan yaitu pelabelan kata harus dilakukan oleh orang yang ahli pada bidang kebahasaan dan tidak seluruh pelabelan kata tepat. Pada penelitian [5] yang dilakukan oleh F. Koto

dan G. Y. Rahmaningtyas, penelitian tersebut membandingkan pelabelan kata yang dilakukan secara manual menggunakan TF dan TF-IDF dengan pelabelan kata secara otomatis menggunakan InSet Lexicon, Vania Lexicon, SentiWordNet, Liu Lexicon, dan AFINN dengan data berbahasa Indonesia. Hasil penelitian ini menunjukkan pelabelan kata secara otomatis menggunakan InSet Lexicon memiliki nilai akurasi tertinggi sebesar 65,78% dibandingkan dengan pelabelan kata secara manual menggunakan TF-IDF yang mendapatkan nilai akurasi sebesar 59,52%.

Berdasarkan penelitian yang telah disebutkan, dapat disimpulkan pelabelan secara manual kurang efektif karena harus dilakukan oleh orang yang ahli dalam bidang kebahasaan dan tidak seluruh label tepat serta membutuhkan waktu yang lebih banyak. Selain itu pelabelan secara otomatis dapat menghasilkan nilai akurasi yang lebih tinggi dibandingkan dengan pelabelan secara manual. Dalam penelitian analisis sentimen menerapkan pelabelan otomatis InSet Lexicon hanya ada beberapa penelitian dan penelitian tersebut belum menerapkan analisis sentimen berdasarkan aspek.

Berdasarkan uraian yang sudah dipaparkan, penelitian ini akan menerapkan pelabelan otomatis *InSet Lexicon* pada analisis sentimen berdasarkan aspek dengan dataset berbahasa Indonesia.

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang yang telah disampaikan, maka dapat diidentifikasi suatu masalah yaitu pelabelan kata secara manual kurang efektif karena harus dilakukan oleh orang yang ahli pada bidang kebahasaan dan tidak seluruh label tepat, ada beberapa kesalahan dalam pemberian label serta membutuhkan waktu yang lama. Selain itu pelabelan kata secara otomatis pada analisis sentimen memiliki tingkat akurasi lebih tinggi dibandingkan pelabelan kata secara manual.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk mengetahui penerapan pelabelan otomatis pada analisis sentimen berdasarkan aspek. Adapun tujuan penelitian ini

adalah untuk mengukur nilai akurasi pada penerapan metode pelabelan otomatis pada analisis sentimen berdasarkan aspek.

1.4 Batasan Masalah

Batasan masalah bertujuan untuk memperkecil cakupan penelitian agar penelitian menjadi terfokus pada permasalahan yang ada. Adapun batasan-batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data yang digunakan pada penelitian ini didapat dari Twitter dengan tag covid dan omicron dengan periode pengambilan data dari tanggal 27 Januari 2022 sampai 31 Januari 2022.
2. Data yang digunakan adalah teks berbahasa Indonesia.
3. Penentuan aspek secara otomatis menggunakan Latent Dirichlet Allocation.
4. *Stopword removal* dan *Stemming* menggunakan library Sastrawi.

1.5 Metode Penelitian

Dalam penelitian ini, tahap penelitian yang akan dilakukan memiliki alur seperti pada Gambar 1.1.

Gambar 1.1 Alur Penelitian

1.5.1 Tahap Perencanaan

Tahap perencanaan dalam penelitian ini memiliki beberapa tahapan, yaitu:

1. Identifikasi Masalah

   Tahapan identifikasi masalah dilakukan dengan memaparkan beberapa penelitian sebelumnya yang berkaitan dengan kasus analisis sentimen dan analisis sentimen berdasarkan aspek, sehingga didapat rumusan masalah mengenai belum diketahuinya apakah pelabelan otomatis memiliki pengaruh pada analisis sentimen berdasarkan aspek.

2. Menetapkan Tujuan Penelitian

Tahapan ini menetapkan apa yang menjadi tujuan dari penelitian ini, yaitu untuk mengukur nilai akurasi pada penerapan metode pelabelan otomatis pada analisis sentimen berdasarkan aspek.

3. Studi Literatur

   Tahapan studi literatur dilakukan dengan cara membaca jurnal ilmiah, buku-buku, dan artikel yang terkait dengan analisis sentimen berdasarkan aspek, metode preprocessing, metode ekstraksi aspek, dan metode pelabelan otomatis.

4. Pengumpulan Dataset

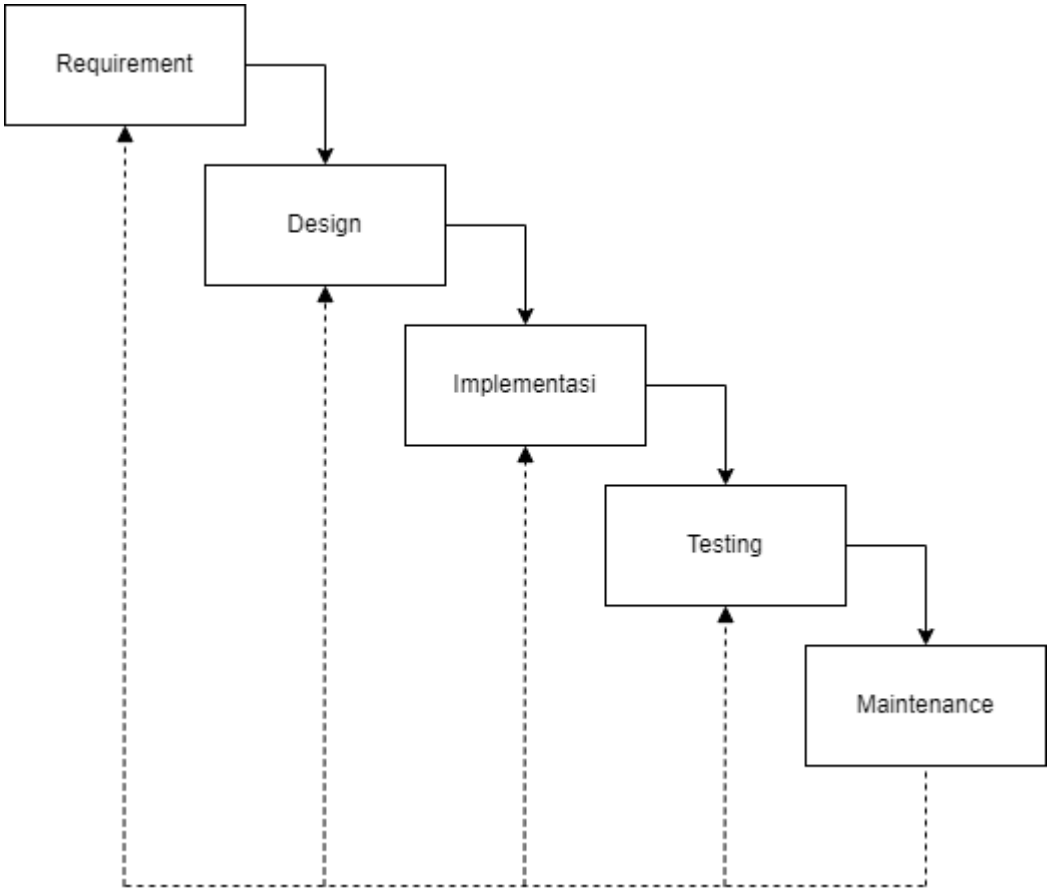
   Dataset yang digunakan pada penelitian ini diambil dari tweet twitter dengan tag omicron.

1.5.2 Tahap Analisis

Tahap analisis dalam penelitian ini akan menganalisis data masukan, analisis preprocessing, analisis metode ekstraksi aspek menggunakan metode LDA, dan analisis metode pelabelan otomatis.

1.5.3 Tahap Pembangunan Perangkat Lunak

Pada tahap ini dilakukan pembangunan perangkat lunak untuk mengetahui hasil sentimen dan nilai akurasi pada analisis sentimen berdasarkan aspek dengan pelabelan kata secara otomatis. Metode yang digunakan dalam membangun perangkat lunak yaitu metode *Waterfall*. Metode *Waterfall* merupakan salah satu metode pengembangan perangkat lunak yang tahapannya berurutan dan sistematis [6]. Berikut adalah tahapan pada metode *Waterfall*.



Gambar 1.2 Tahapan Waterfall

1.5.3.1 Requirement

Requirement merupakan tahapan yang dilakukan untuk mempersiapkan dan menganalisis kebutuhan dari suatu aplikasi. Tahapan ini bertujuan untuk mengetahui masalah dan setiap batasan dari aplikasi.

1.5.3.2 Design

Design merupakan tahapan yang dilakukan untuk mendesain aplikasi sesuai dengan perencanaan yang telah dibuat. Pada tahapan ini akan menghasilkan alur sistem dan gambaran jelas mengenai tampilan antarmuka aplikasi.

1.5.3.3 Implementasi

Implementasi merupakan tahapan yang dilakukan untuk merealisasi desain dan rancangan yang telah dibuat menggunakan tools dan bahasa pemograman serta mengimplementasi algoritma yang sesuai dengan penelitian ini.

1.5.3.4 Testing

Testing merupakan tahapan pengujian aplikasi. Pada tahapan ini semua fungsi dan fitur akan diuji serta dievaluasi.

1.5.3.5 Maintenance

Maintenance merupakan tahapan pemeliharaan aplikasi yang sudah jadi dan siap dioperasikan oleh pengguna. Pemeliharaan aplikasi memeungkinkan pengembang untuk melakuka perawatan dan perbaikan apabila ditemukan kesalahan yang tidak terdeteksi pada tahapan sebelumnya.

1.5.4 Tahap Pengujian

Pada tahap ini dilakukan pengujian terhadap hasil sentimen yang sudah didapat untuk mengetahui seberapa besar nilai akurasi yang dihasilkan oleh metode pelabelan otomatis yang diterapkan pada analisis sentimen berdasarkan aspek.

1.5.5 Tahap Penarikan Kesimpulan

Pada tahap ini dilakukan penarikan kesimpulan berdasarkan hasil dari tahap pengujian mengenai penerapan pelabelan otomatis dalam analisis sentimen berdasarkan aspek.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini disusun untuk memberikan gambaran secara umum mengenai penelitian yang akan dilakukan agar hasil dari laporan penelitian ini lebih terarah dan tidak menyimpang. Sistematika penulisan pada penelitian ini adalah sebagai.

BAB 1 PENDAHULUAN

Bab ini membahas tentang latar belakang permasalahan, identifikasi masalah, maksud dan tujuan, batasan masalah penelitian, metode penelitian yang digunakan dan sistematika penulisan dari penelitian mengenai penerapan pelabelan otomatis pada analisis sentimen berdasarkan aspek.

BAB 2 LANDASAN TEORI

Bab ini membahas berbagai konsep dasar dan teori-teori dari para ahli yang berkaitan dengan topik penelitian seperti analisis sentimen, analisis sentimen

berdasarkan aspek, *preprocessing*, pelabelan data, *Latent Dirichlet Allocation* (LDA), dan *Confusion Matrix* serta dapat mendukung pokok permasalahan tentang penerapan pelabelan otomatis pada analisis sentimen berdasarkan aspek.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini membahas analisis masalah, analisis data, analisis sistem, analisis data inputan, analisis tahapan *preprocessing*, analisis ekstraksi aspek menggunakan model LDA, analisis pelabelan kata secara otomatis menggunakan *Inset Lexicon*, analisis klasifikasi sentimen menggunakan algoritma Naïve Bayes, analisis kebutuhan non fungsional dan analisis kebutuhan fungsional.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang penerapan atau implementasi dari tahap analisis serta pengujian hasil dari perangkat lunak yang telah dibangun apakah perangkat lunak sesuai dengan yang diharapkan.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan hasil dari peneltian penerapan pelabelan otomatis pada analisis sentimen berdasarkan aspek yang telah dilakukan dan nilai akurasi yang didapat serta saran-saran yang dapat diberikan untuk penelitian selanjutnya.