

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Seiring dengan perkembangan internet sebagai media daring di mana jutaan orang dapat mengekspresikan opininya secara bebas dalam berbagai ragam topik telah menjadikan internet sebagai sumber informasi yang tidak terbatas [1]. Dengan berkembang pesatnya pengguna internet, semakin berkembang pula media sosial. Media sosial merupakan media daring, dimana penggunanya bisa mengakses informasi, berpartisipasi, berbagi dan menciptakan konten. Salah satu jejaring media sosial yang marak digunakan adalah *twitter*. Melalui *twitter* pengguna dapat mengekspresikan tulisannya dalam 140 karakter. Saat ini terdapat 330 juta pengguna aktif bulanan dan 145 juta pengguna aktif harian di *twitter* dimana ada 500 juta *tweet* dikirim setiap hari [2].

Kelebihan *twitter* dibanding media sosial lain diantaranya adalah jangkauannya yang luas, potensi periklanan di masa mendatang yang lebih besar, komunikasi yang terjadi sangat cepat dan lebih terukur dari media sosial lain. *Twitter* membantu penyebaran informasi secara lebih cepat dan menjadikan informasi tersebut menjadi sebuah topik yang dibahas oleh para penggunanya. Saat ini banyak media massa dan *brand* yang menggunakan *twitter* sebagai penyampai informasi [3]. *Twitter* juga menjadi media untuk mengakses pendapat dan pemikiran penting dari orang yang berbeda tentang berbagai masalah bisnis, politik, kesehatan dan sosial yang menarik banyak perhatian untuk memperkirakan sentiment pengguna di *twitter*[4].

Tantangan dari data *tweet* adalah luasnya topik yang dibahas, dapat dikatakan bahwa orang-orang membuat *tweet* tentang apa saja dengan cara yang berbeda satu sama lain. Selain itu, tantangan dari data *tweet* adalah data *tweet* memiliki segmen yang sangat pendek dimana informasi pada data *tweet* kurang jelas didapatkan dan juga pada data *tweet* terdapat banyak objek dan emosi sentimen yang menjadi tantangan pada analisis sentimen pada data *twitter* [5].

Pada penelitian ini, terdapat target dan konteks dari dataset *twitter* yang akan diproses. Pengertian dari konteks adalah suatu uraian atau kalimat [6], dimana

konteks pada penelitian ini adalah *tweet* dari pengguna *twitter*. Sedangkan target adalah sasaran yang sudah ditetapkan untuk dicapai [7], *tweet* dari pengguna *twitter* diklasifikasikan sentimennya terhadap target yang diberikan yaitu positif, negatif dan netral. Pengguna mungkin menyebutkan beberapa target dalam satu *tweet*. Misalnya, *tweet* “@tim\_cook windows better than mac” pada *tweet* ini terdapat tiga target (@tim\_cook, windows, dan mac”). Pengguna masing-masing mengekspresikan sentimen netral, positif, dan negatif secara berturut-turut. Jika informasi target diabaikan, sulit untuk mendapatkan sentimen yang benar untuk target tersebut.

Pendekatan-pendekatan tradisional yang salah satunya dilakukan oleh Jiang dkk. [8] yang berfokus pada perancangan dengan serangkaian fitur seperti *bag-of-word* dan *sentiment lexicon* untuk melatih *classifier* (mis., SVM) yang digunakan pada klasifikasi sentimen berbasis target terdapat tantangan pada *feature engineering* yang cukup sulit dan menjadi *bottleneck* pada performa. Selain itu, Jiang dkk. [8] menunjukkan bahwa sekitar 40 % kesalahan pada klasifikasi sentimen disebabkan oleh tidak mempertimbangkan target dalam proses klasifikasi yang membuat banyak peneliti mengembangkan teknik *deep learning* dengan merancang *neural network* yang efektif untuk secara otomatis menghasilkan representasi *low dimensional* dari target dan konteks [9] [10] [11] [12].

Pada penelitian terkait diatas dapat dikatakan pentingnya target pada pengembangan berbagai metode dengan tujuan memodelkan konteks secara tepat melalui representasi *target-specific*. Namun, penelitian terkait diatas mengabaikan permodelan target yang terpisah, terutama dengan bantuan konteks. Menurut Ma dkk. [9] koordinasi target dan konteks dapat meningkatkan kinerja klasifikasi sentimen berbasis target. Misalnya, “@billgates windows make my battery life too short”. Ketika “short” pada “windows make my battery life too short”, sentimennya menjadi negatif. Tetapi, ketika “short” digunakan pada *tweet* “@john\_mayer short fat noodle spoon”, sentimennya menjadi netral. Berdasarkan pentingnya koordinasi target dan konteks untuk dapat meningkatkan kinerja klasifikasi sentimen, penelitian ini akan menggunakan model *interactive attention network* (IAN) yang diusulkan oleh Ma dkk [9] yang digunakan pada data *tweet* pengguna *twitter*. Ma dkk. [9] mengusulkan model IAN yang berbasis pada *long short term memory*

(LSTM) dan *attention mechanism*. IAN menggunakan *attention mechanism* yang berasosiasi dengan target untuk mendapatkan informasi penting dari konteks dan menghitung representasi konteks untuk klasifikasi sentimen. Lebih lanjut, IAN menggunakan informasi interaktif dari konteks untuk men-*supervise* pemodelan target yang berguna untuk menilai sentimen. Dengan representasi target dan representasi konteks yang digabungkan, IAN memprediksi polaritas sentimen untuk target dalam konteksnya. Berdasarkan eksperimen SemEval 2014 menunjukkan IAN dapat secara tepat memodelkan target dan konteks, dan mencapai performa *state-of-the-art*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana *Interactive Attention Network* memberikan model untuk memprediksi polaritas sentimen berbasis target pada dataset *twitter*?
2. Seberapa penting koordinasi target dan konteks untuk meningkatkan kinerja klasifikasi berbasis target?

## 1.3 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah:

1. Menghasilkan model untuk memprediksi polaritas sentimen berbasis target pada dataset *twitter* menggunakan *Interactive Attention Network*.
2. Membuktikan pentingnya koordinasi target dan konteks untuk meningkatkan kinerja klasifikasi berbasis target.

## 1.4 Batasan Masalah

Pada penelitian ini terdapat beberapa batasan masalah yang perlu diperhatikan, yaitu:

1. Dokumen yang digunakan adalah *tweet* pengguna *twitter* berbentuk teks.
2. Proses klasifikasi sentimen dilakukan pada *tweet* berbahasa Inggris.
3. Menggunakan data *twitter* yang digunakan pada penelitian Dong dkk. [10].
4. Target dan konteks sudah diketahui pada dataset.
5. Keluaran dari model adalah polaritas sentimen.

## **1.5 Metodologi**

Terdapat beberapa metodologi yang penulis gunakan dalam melakukan penelitian dan penyusunan tugas akhir ini, yaitu sebagai berikut :

1. Studi Literatur

Studi literatur adalah metode dalam pengumpulan data yang berkaitan dengan penelitian terkait. Data yang dikumpulkan berasal dari buku yang terkait dengan keilmuan yang akan diterapkan serta dari jurnal – jurnal penelitian terkait yang telah dilakukan sebelumnya.

2. Analisis

Pada tahap ini dilakukan analisis terhadap studi literatur untuk memahami permasalahan, kebutuhan yang diperlukan untuk tahap perancangan, dan tahapan yang perlu dilakukan untuk menerapkan metode yang dapat menyelesaikan permasalahan sesuai dengan tujuan penelitian.

3. Perancangan

Pada tahap ini, rancangan penelitian didesain berdasarkan hasil analisis. Rancangan penelitian meliputi rancangan data, arsitektur, dan pelatihan

4. Implementasi

Implementasi dilakukan dengan menerapkan hasil analisis dan perancangan yang sudah dilakukan pada tahapan sebelumnya,

5. Pengujian

Pengujian dilakukan untuk menguji model implementasi dengan mengujinya pada himpunan data tes. Dengan begitu, hasil akurasi dari model yang diimplementasikan dengan model lain.

## **1.6 Sistematika Penulisan**

Sistematika penulisan penelitian terdiri dari 5 bab. Setiap bab terdiri dari sub – sub bab dan penjelasan yang tersusun agar mudah dipahami. Berikut penjelasan tentang masing – masing bab tersebut :

**Bab I           Pendahuluan**

Pada bab ini berisikan deskripsi umum dari penelitian ini. Bab ini terdiri dari sub-bab latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi dan sistematika penulisan.

**Bab II           Studi Literatur**

Pada bab ini berisikan teori – teori dan penelitian yang terkait dengan penelitian ini.

**Bab III          Metodologi Penelitian**

Pada bab ini berisikan pedoman tentang tahapan – tahapan yang dilakukan pada penelitian ini agar hasil dari penelitian ini tidak menyimpang dari tujuan penelitian yang telah ditentukan sebelumnya.

**Bab IV          Hasil dan Pembahasan**

Pada bab ini berisikan penjelasan mengenai hasil dari tahapan – tahapan penelitian yang telah dilakukan sebelumnya. Pada bab ini juga memberikan pembahasan mengenai hasil tersebut, apakah sudah berjalan dengan sesuai yang diharapkan atau tidak.

**Bab V           Simpulan dan Saran**

Pada bab ini berisikan kesimpulan dan saran yang berkaitan dengan penelitian ini.