

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Pada era digitalisasi ini, dengan akses global ke internet, media sosial menjadi sumber utama dalam menyampaikan aspirasi dan pendapat secara daring. Di antara berbagai *platform* media sosial, Twitter telah menjadi salah satu tempat utama orang-orang menyampaikan aspirasi, pendapat, dan pandangan mereka melalui cuitan atau *tweet* [1]. Setiap harinya, diperkirakan terdapat 500 juta *tweet* pada media sosial Twitter [2]. Fenomena tersebut menggambarkan tingginya interaksi pengguna, penyampaian opini, serta pertukaran informasi melalui platform Twitter dibanding dengan sosial media lain yang lebih fokus kepada foto dan video [3].

Online reviews atau ulasan daring berperan sangat penting dalam membantu pengambilan keputusan pembelian sebuah produk [4]. Konsumen percaya bahwa ulasan daring dapat membantu mereka untuk meningkatkan akurasi keputusan pembelian, mengurangi resiko kerugian, dan memengaruhi pilihan belanja mereka [5]. Sementara itu, konten ulasan daring memiliki volume yang tinggi, kecepatan tinggi, dan kebenaran yang tinggi. Oleh karena itu, ulasan daring dapat diklasifikasikan sebagai "Big Data" dan harus dianalisis dengan pendekatan analitik yang sesuai [5] [6]. Atas dasar tersebut, saat ini terdapat aplikasi analisis sosial media untuk menganalisis "Big Data" pada media sosial Twitter yang memiliki fitur untuk mengidentifikasi akun-akun berpengaruh (*buzzer*) dan menganalisis topik-topik yang dibahas dalam sebuah topik yang lebih besar. Aplikasi ini dapat digunakan secara umum, termasuk konsumen yang ingin membeli produk. Melalui fitur "Topic Modelling" konsumen dapat mengetahui topik apa saja yang dibahas di sekumpulan *tweet* yang berkaitan dengan topik ulasan produk sebagai wawasan untuk mengenal produk yang akan dibeli. Namun, informasi yang disampaikan pada fitur tersebut belum sepenuhnya memenuhi kebutuhan konsumen itu sendiri (LAMPIRAN A)

Para konsumen memerlukan informasi berupa kesimpulan apakah ulasan produk tersebut bersifat positif atau negatif, yang mana

kesimpulan tersebut dapat memudahkan mereka dalam mengambil sebuah keputusan salah satunya keputusan dalam pembelian produk Error! Reference source not found.. Kebutuhan informasi ini berdasarkan survei yang menunjukkan bahwa konsumen sering mencari ulasan daring di Twitter sebagai pertimbangan utama dalam keputusan pembelian Error! Reference source not found.. Ulasan Twitter memberikan informasi mendalam tentang pengalaman orang lain, membantu konsumen menilai apakah produk layak dibeli. Banyak ulasan positif meningkatkan kemungkinan pembelian, sedangkan ulasan negatif mengurangnya [7]. Namun, dalam praktik saat ini, konsumen sering kesulitan menyimpulkan sentimen mayoritas dari ulasan karena sulit mengingat dan menilai semua ulasan, yang menyebabkan kebingungan dan kesalahan dalam keputusan akhir. Meskipun demikian, konsumen tetap membeli produk dengan cara menelusuri ulasan *tweet* secara satu per satu. Namun, mereka sering merasa khawatir dan kecewa terhadap produk yang dibeli karena kurangnya informasi mengenai mayoritas ulasan, apakah lebih banyak yang positif atau negatif (LAMPIRAN A).

Dengan demikian, diperlukan pengembangan aplikasi dan metode untuk menganalisis topik pada Twitter salah satunya terkait topik ulasan produk, sehingga analisis tersebut dapat membantu pengguna aplikasi termasuk konsumen dalam mendapatkan informasi berupa kesimpulan yang jelas mengenai apakah suatu produk memiliki ulasan yang lebih banyak positif atau negatif, sehingga memudahkan mereka dalam mengambil keputusan pembelian. Pendekatan berbasis analisis sentimen yang dapat menjadi solusi untuk tantangan ini [8]. Metode *deep learning* telah terbukti lebih unggul dalam mengklasifikasikan sentimen teks [9] [10]. *Deep learning*, khususnya dengan penggunaan *neural networks*, mampu mengenali pola-pola kompleks dan korelasi dalam data, yang berkontribusi pada akurasi yang lebih tinggi dibandingkan dengan teknik *machine learning* konvensional [11]. Algoritma CNN merupakan algoritma *deep learning* yang sangat baik dalam mengenali dan mengekstraksi fitur lokal dari data input, seperti pola kata atau frasa dalam teks. Kim dan Yoon [12] menggunakan CNN untuk klasifikasi teks karena operasi konvolusi bisa mendapatkan fitur tingkat tinggi dari kata-kata lokal. Namun, hanya menggunakan CNN tidak memungkinkan untuk

mendapatkan informasi posisi dan urutan, yang penting untuk analisis sentimen teks yang lebih panjang [13]. Disisi lain, algoritma *hybrid* yang menggabungkan Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) terbukti dapat bekerja dengan baik juga dikarenakan dengan menggabungkan kedua algoritma tersebut, CNN mampu menangkap pola-pola penting dalam teks seperti frase, sementara LSTM dapat memahami urutan dan konteks temporal dari kata-kata tersebut [13] [14]. Penelitian sebelumnya menunjukkan bahwa penggabungan CNN-LSTM unggul dalam klasifikasi sentimen dibandingkan dengan menggunakan satu model algoritma saja pada teks yang relatif panjang. Hal ini menarik untuk diteliti lebih lanjut, terutama apakah kombinasi CNN-LSTM ini akan menghasilkan perbedaan yang signifikan dalam analisis teks pendek seperti data Twitter, serta apakah ada perubahan dalam pengetahuan yang dihasilkan setelah penggabungan layer tersebut.

1.2 Rumusan Masalah

Dari uraian latar belakang di atas, dapat diuraikan bahwa rumusan masalahnya adalah apakah dengan penerapan analisis sentimen dapat mengklasifikasikan sentimen terkait ulasan sebuah produk ?

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah mengembangkan aplikasi analisis sosial media dengan menerapkan analisis sentimen untuk mengklasifikasikan *tweet* yang sedang dibicarakan pada Twitter salah satunya topik ulasan produk. Adapun tujuan penelitian ini adalah sebagai berikut:

1. Memberikan pengetahuan bagi pengguna aplikasi analisis sosial media untuk pengambilan keputusan sesuai dengan kebutuhan pengguna. Salah satunya konsumen dalam mendapatkan informasi berupa kesimpulan apakah sebuah produk memiliki sentimen yang positif atau negatif, sehingga dapat membantu pengambilan keputusan pembelian produk tersebut.
2. Membandingkan pengetahuan yang dihasilkan dari kedua metode yaitu CNN dan CNN-LSTM terhadap pengguna aplikasi analisis sosial media,

sehingga bisa mengetahui apakah ada perbedaan pengetahuan yang dihasilkan dari penggabungan algoritma CNN-LSTM.

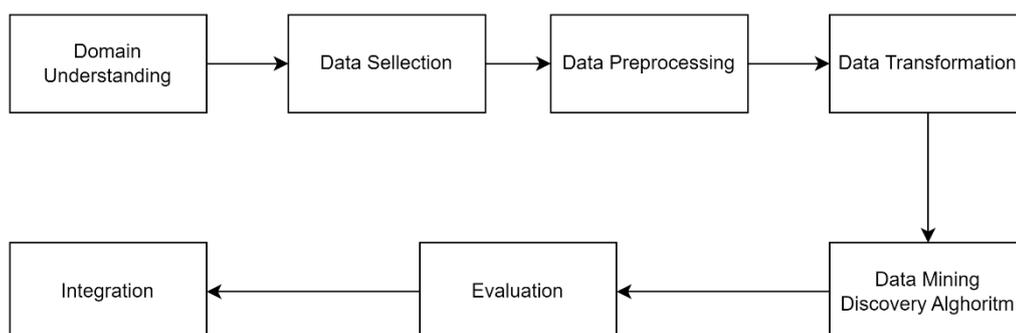
1.4 Batasan Masalah

Adapun untuk batasan masalah pada pelaksanaan skripsi ini adalah sebagai berikut.

1. Data tweet yang diambil berbahasa Indonesia.
2. Algoritma yang digunakan adalah CNN dan CNN-LSTM
3. Pengumpulan data tweet menggunakan *service* topic modelling pada aplikasi analisis sosial media.
4. Kelas sentimen yang digunakan terdiri dari positif dan negatif [15] [16] [17].

1.5 Metodologi Penelitian

Penelitian ini menggunakan metode penelitian Knowledge Discovery in Database [18] yang kemudian disesuaikan dengan dengan penelitian yang dilakukan sehingga menghasilkan sebagaimana yang dijelaskan pada **Gambar 1.1 Metodologi Penelitian**.



Gambar 1.1 Metodologi Penelitian

Berikut adalah penjelasan dari masing-masing tahapan :

1) Domain Understanding

Merupakan langkah awal untuk memecahkan masalah. Pengumpulan dan pemahaman informasi dilakukan untuk menentukan

fungsionalitas dari pengetahuan yang dihasilkan dan nilai apa yang diberikannya kepada pengguna.

2) Data Selection

Tahapan untuk memilih dan membuat dataset tempat discovery akan dilakukan, termasuk menemukan data yang sudah tersedia dan memperoleh data tambahan yang diperlukan, lalu mengintegrasikan semua data tersebut pada proses KDD pada satu dataset, termasuk atribut yang nanti akan digunakan.

3) Data Preprocessing

Merupakan proses yang bertujuan untuk membersihkan dan menyiapkan data agar siap untuk dianalisis. Langkah-langkah *preprocessing* ini terdiri dari seleksi atribut, penambahan atribut, penggabungan data, pembersihan data, dan pemotongan dataset.

4) Data Transformation

Data transformation adalah tahap yang melibatkan perubahan format atau struktur data agar lebih sesuai untuk analisis. Proses ini bertujuan untuk mengubah data mentah menjadi format yang dapat digunakan oleh algoritma analitik. Hal ini meliputi *feature extraction* yaitu mengambil fitur-fitur yang relevan dari data mentah. Dalam analisis teks, ini bisa berupa ekstraksi n-gram, pembobotan kata menggunakan TF-IDF, atau pembentukan embedding kata dengan teknik seperti Word2Vec.

5) Data Mining Discovery Algoritm

Merupakan tahapan penerapan algoritma deep learning CNN dan CNN-LSTM untuk menemukan pengetahuan apakah topik yang sedang dibahas tersebut memiliki sentimen positif atau negatif lalu membuat model klasifikasi nya..

6) Evaluation

Merupakan pengujian sistem secara keseluruhan pada percobaan pengguna, untuk mengetahui apakah sistem dapat menghasilkan menghasilkan output yang sesuai dan memastikan bahwa sistem dapat menghasilkan wawasan bagi pengguna

7) Integration

Tahap integration yaitu pembuatan sistem secara keseluruhan dengan mengimplementasikan hasil analisis ke dalam perangkat lunak untuk memberikan insight tentang sentiment yang ditemukan. Pembangunan perangkat lunak yang dibuat berupa sistem berbasis *website*.

1.6 Sistematika Penelitian

Sistematika penulisan ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan pada laporan ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Pada bab ini menjelaskan mengenai latar belakang masalah, perumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian yang digunakan, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Pada bab ini, merupakan penjelasan tentang teori dasar yang akan digunakan dalam penelitian ini. Teori-teori ini termasuk teori yang relevan dan termasuk dalam topik penelitian ini, seperti analisis sentimen, *text preprocessing*, dan *artificial neural network/deep learning* dll.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Pada bab ini akan dibahas secara rinci hasil analisis objek penelitian yang mencakup identifikasi masalah yang muncul serta kebutuhan sistem yang dibutuhkan untuk memecahkan masalah tersebut.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Pada bab ini akan dijelaskan bagaimana hasil analisis digunakan. Selain itu, akan dijelaskan bagaimana teknik penanganan digunakan, diuji, dan disesuaikan dengan hasil penelitian.

BAB 5 KESIMPULAN DAN SARAN

Pada bab terakhir ini menjelaskan mengenai kesimpulan yang didapatkan dari hasil penelitian ini dan saran untuk pengembangan ke depannya.