

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Bisnis kuliner di Indonesia memiliki prospek bisnis yang baik dengan didukung oleh perkembangan jaman yang semakin maju. Tujuan dari suatu bisnis kuliner adalah meningkatkan dan memaksimalkan pemenuhan kebutuhan pelanggan. Agar bisnis kuliner tetap bertahan dan maju, tujuan dari bisnis tersebut harus terpenuhi dengan memberikan kenyamanan dan kepuasan pada pelanggan. Pendapat pelanggan juga sangat penting agar bisnis dapat meningkatkan kualitas bisnisnya, baik dari segi pelayanan maupun fasilitas. Pendapat online juga telah menjadi faktor penting untuk pelanggan, bahkan saat ini ulasan berfungsi sebagai alat pengambilan keputusan utama bagi pelanggan dalam membeli beberapa produk [1]. Maka dari itu, perlu untuk mendapatkan data opini pelanggan dan melakukan Analisis Sentimen terhadap data yang berisi opini pelanggan tersebut untuk dapat mengetahui perspektif pelanggan terhadap aspek pada restoran.

Analisis Sentimen, atau yang sering disebut juga dengan Opinion Mining adalah ilmu yang menganalisis pendapat seseorang, sentimen, perilaku, penilaian dan emosi terhadap suatu entitas, seperti produk, pelayanan, jasa, organisasi, individu, dan entitas lainnya. Analisis sentimen ini dapat mengklasifikasikan polaritas teks dari sebuah kalimat atau dokumen untuk mengetahui apakah opini dari kalimat atau dokumen tersebut memiliki polaritas positif atau negatif [2]. Pada analisis sentimen terdapat beberapa level, seperti level dokumen, level kalimat, dan level aspek. Analisis sentimen level aspek akan menunjukkan performa yang lebih baik dibandingkan level dokumen dan level kalimat [3]. Hal ini dikarenakan analisis sentimen level aspek menganalisis teks atau kalimat secara lengkap, yaitu menentukan sentimen dari setiap aspek yang ada pada suatu kalimat. Oleh karena itu, analisis yang perlu dilakukan adalah analisis sentimen berbasis aspek.

Pada penelitian terdahulu oleh M.I. Ali . telah melakukan penelitian dengan menerapkan Analisis Sentimen (Opinion Mining) pada tingkat aspek menggunakan metode Elman Recurrent Neural Network. Dataset yang digunakan adalah review restoran dalam bahasa Indonesia, dan aspek yang dipakai adalah aspek Food, aspek Price, aspek Service dan aspek Ambience. Hasil penelitian tersebut menunjukkan pengujian terhadap hasil klasifikasi sentimen dengan F1 score sebesar 89% untuk label "bukan sentimen", 21% untuk label "negatif", dan 56% untuk label "positif". Nilai weighted F1 score secara keseluruhan adalah 82,27% dan rata-rata akurasi yang didapatkan sebesar 81,22%. Proses yang dilakukan untuk mengubah kata ke dalam bentuk vektor yaitu dengan one-hot encoding, dimana diharuskan membangun kamus kata terlebih dahulu dan tidak jarang ditemukannya dalam kalimat token yang tidak terdapat pada kamus kata, serta dimensi vektor kata yang dihasilkan sangat besar [4].

Penelitian yang dilakukan oleh P. Liu, S. Joty dan H. Meng membandingkan beberapa arsitektur RNN dengan menggunakan beberapa jenis word embedding. Akurasi yang didapatkan dengan menerapkan SENNA embeddings pada arsitektur RNN adalah 78.83% untuk Jordan Recurrent Neural Network, 81.36% untuk Elman Recurrent Neural Network, dan 79.43% untuk LSTM. Dari hasil penelitian tersebut, maka arsitektur Elman dengan menggunakan word embedding memiliki akurasi yang tinggi dari arsitektur RNN lainnya [5]. N. M. Alharbi, N. S. Alghamdi, E. H. Alkhamash dan J. F. Al Amri dalam penelitiannya menggunakan beberapa jenis word embedding, yaitu Glove, word2vec dan FastText. Tahap preprocessing yang digunakan antara lain case folding, spelling correction, tokenization, stop words, removed punctuation, dan lemmatization, setelah preprocessing, data akan diubah menjadi bentuk vektor dengan word embedding yang dilakukan tanpa membutuhkan kamus data maupun proses one-hot-encoding. Hasil dari penelitian tersebut menunjukkan penggunaan FastText pada algoritma RNN menghasilkan akurasi terbaik sebesar 93.75% untuk dataset yang tidak seimbang dan sebesar 88.39% untuk dataset yang seimbang [6].

Word embedding adalah salah satu pendekatan distributed word representation dan merupakan teknik pembelajaran fitur pada Natural Language Processing (NLP) yang mengubah kata menjadi angka dalam bentuk vektor dengan dimensi tertentu. Hasil *embedding* digunakan dalam pemodelan *machine learning* dan berfungsi sebagai dasar untuk komputasi NLP. Pendekatan *word embedding* lebih kuat dibandingkan dengan *one-hot encoding*, karena setiap kata yang memiliki makna yang sama akan memiliki pengkodean yang hampir serupa [7].

Berdasarkan latar belakang yang telah dijelaskan dan beberapa penelitian yang sudah dilakukan sebelumnya, maka studi kasus pada penelitian ini adalah Analisis Sentimen Berbasis Aspek terhadap opini bahasa Indonesia yang diberikan oleh pelanggan tentang restoran. Analisis Sentimen dilakukan pada level aspek dengan menggunakan metode Elman Recurrent Neural Network dengan menggunakan *word embedding* sebelum diklasifikasikan Elman RNN. Hasil dari penelitian ini diharapkan dapat meningkatkan akurasi pada kasus ASBA terhadap opini berbahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan uraian yang terdapat pada latar belakang masalah, maka dapat dirumuskan masalah yang ada yaitu bagaimana pengaruh penggunaan *word embedding* FastText terhadap performansi metode Elman Recurrent Neural Network dalam menganalisis sentimen berbasis aspek terhadap *review* restoran bahasa Indonesia.

1.3 Maksud dan Tujuan

Berdasarkan permasalahan yang telah teridentifikasi, maka maksud dari penelitian ini adalah mengimplementasikan *word embedding* FastText dan metode Elman Recurrent Neural Network pada analisis sentimen berbasis aspek terhadap *review* restoran bahasa Indonesia.

Sedangkan tujuan yang diharapkan dengan dibangunnya sistem ini adalah untuk mengetahui pengaruh penggunaan *word embedding* FastText pada tingkat

akurasi dan nilai F1 Score dengan menggunakan metode Elman Recurrent Neural Network terhadap *review* restoran bahasa Indonesia berdasarkan aspek-aspek terkait.

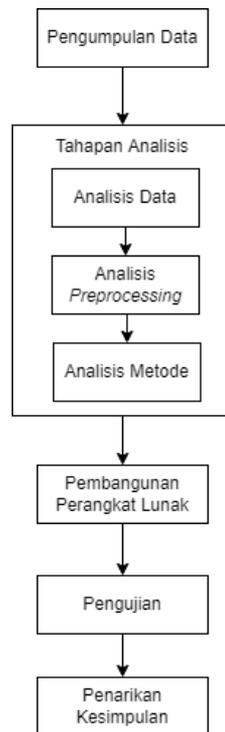
1.4 Batasan Masalah

Adapun batasan masalah pada penelitian yang akan dilakukan adalah sebagai berikut:

- a. Data opini yang digunakan adalah teks berbahasa Indonesia data berasal dari penelitian M.I.Ali.
- b. Tahapan text preprocessing yang dilakukan pada penelitian ini meliputi: *case folding, filtering, word normalization, tokenization, dan stopwords removal.*
- c. Setelah tahapan preprocessing, data latih akan melalui tahapan *word embedding*. *Word embedding* yang akan digunakan adalah FastText.
- d. Aspek yang akan dikenali pada penelitian ini adalah aspek *Food*, aspek *Price*, aspek *Service* dan aspek *Ambience*.
- e. Setiap aspek akan dinilai sentimennya menjadi bukan sentimen, sentimen positif, dan sentimen negatif.

1.5 Metodologi Penelitian

Adapun tahap penelitian yang akan dilakukan dengan alur yang dapat dilihat pada Gambar 1.1.



Gambar 0.1 Alur Penelitian

1.5.1 Pengumpulan Data

Metode pengumpulan data yang digunakan pada penelitian ini adalah sebagai berikut:

1. Studi Literatur, merupakan teknik pengumpulan data dengan cara membaca literatur, jurnal, paper dan sumber bacaan yang mendukung dengan topik penelitian ini. Studi literatur yang dibutuhkan adalah mengenai analisis sentimen, metode Elman Recurrent Neural Network, dan word embedding.
2. Pengumpulan Dataset, dataset yang digunakan pada penelitian ini berupa review bahasa Indonesia pada restoran yang diambil dari penelitian M.I.Ali.

1.5.2 Analisis

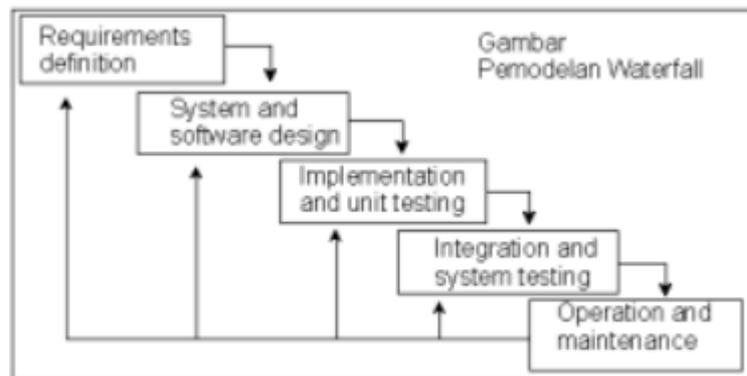
Analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Analisis Data, tahapan ini meliputi analisis pola kalimat dari data yang didapat.

2. Analisis Preprocessing, pada tahap ini dilakukan agar dapat mengolah data menjadi informasi baru agar karakteristik data tersebut menjadi lebih mudah dimengerti dan berguna untuk permasalahan yang berkaitan dengan penelitian. Adapun langkah-langkah yang akan dilakukan pada tahapan preprocessing adalah *case folding*, *filtering*, *word normalization*, *tokenization*, *stopwords removal*, dan *word embedding*.
3. Analisis Metode, pada tahap ini akan dilakukan analisis klasifikasi menggunakan metode Elman Recurrent Neural Network untuk mengetahui karakteristik dan performa dari algoritma Elman Recurrent Neural Network menggunakan *word embedding* pada kasus analisis sentimen berbasis aspek.

1.5.3 Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan dalam penelitian ini adalah model waterfall. Metode Waterfall ini digunakan karena pada penelitian tahap demi tahap yang dilalui harus menunggu selesainya tahap sebelumnya [8]. Tahapan – tahapan pada metode waterfall seperti pada Gambar 1.2.



Gambar 0.2 Metode Waterfall

a. Requirements Definition

Pencarian kebutuhan sistem dan software pada penelitian ini, seperti analisis fungsional dan non-fungsional yang dibutuhkan pada sistem yang akan dibangun.

b. System and Software Design

Proses ini digunakan untuk membuat desain sistem sesuai dengan kebutuhan pada hasil analisis untuk aplikasi yang akan dibangun.

c. Implementation and Unit Testing

Pada tahap ini, hasil desain sistem diimplementasikan ke dalam program perangkat lunak melalui proses coding, sehingga dapat dimengerti oleh mesin atau komputer, serta melakukan penerapan algoritma *Elman Recurrent Neural Network*.

d. Integration and System Testing

Pengujian terhadap aplikasi yang telah dibangun pada tahap *implementation*. Pengujian dilakukan agar mengetahui apakah terdapat error pada sistem dan sistem sudah sesuai dengan kebutuhan.

e. Operation and Maintenance

Ketika aplikasi dijalankan mungkin saja masih ada error kecil yang tidak ditemukan sebelumnya, atau ada penambahan fitur-fitur yang belum ada pada aplikasi, maka pengembangan diperlukan ketika adanya perubahan.

1.5.4 Pengujian

Pengujian dilakukan untuk menghitung seberapa besar nilai akurasi dan nilai *F1 Score* pada tiap kelas sentimen dalam pengimplementasian metode Elman Recurrent Neural Network menggunakan *word embedding* pada analisis sentimen berbasis aspek terhadap *review* bahasa Indonesia pada restoran dengan metode confusion matrix.

1.5.5 Penarikan Kesimpulan

Pada tahap ini akan dilakukan penarikan kesimpulan berdasarkan hasil dari tahap pengujian dalam penerapan metode Elman Recurrent Neural Network pada kasus analisis sentimen berdasarkan aspek.

1.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran umum tentang penulisan tugas akhir yang akan dilakukan. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Pada Bab 1 akan dijelaskan tentang latar belakang masalah yang diambil, identifikasi masalah, maksud dan tujuan, batasan masalah, metodologi penelitian dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada Bab 2 akan dijelaskan tentang berbagai konsep dasar mengenai analisis sentimen, Elman Recurrent Neural Network, dan teori-teori pendukung lainnya yang berkaitan dengan topik pembangunan perangkat lunak.

BAB 3 ANALISIS DAN PERANCANGAN

Pada Bab 3 ini berisi tentang analisis dan perancangan aplikasi yang dibangun yang meliputi analisis masalah, analisis data masukan, analisis sistem, analisis preprocessing, analisis algoritma Elman Recurrent Neural Network, analisis non fungsional dan fungsional.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Pada Bab 4 ini berisi tentang gambaran dari hasil implementasi dan pengujian pada sistem yang mulai pada proses pengklasifikasian oleh metode Elman Recurrent Neural Network dan melakukan tahapan pengujian dengan metode pengujian yang akan menampilkan hasil keluaran seperti nilai recall, precision dan F1 score pada sistem aplikasi yang telah diprogram untuk menguji akurasi dari analisis sentiment berdasarkan aspek terhadap review restoran berbahasa Indonesia.

BAB 5 KESIMPULAN DAN SARAN

Pada Bab 5 akan dijelaskan tentang kesimpulan yang sudah diperoleh dari hasil penulisan tugas akhir dan saran agar dapat dilakukan pengembangan untuk penelitian selanjutnya.