

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Penelitian

Bahasa Indonesia merupakan bahasa resmi nasional, identitas bangsa dan lambang kebanggaan nasional, merupakan alat penghubung antardaerah dan antarbudaya, serta bahasa pemersatu berbagai masyarakat Indonesia yang memiliki latar belakang sosial budaya yang berbeda-beda (semenjak diikrarkannya Sumpah Pemuda pada 28 Oktober 1928). Mengingat fakta-fakta tersebut, sebagai bangsa Indonesia, sudah seharusnya kita menjunjung tinggi bahasa Indonesia dengan cara menggunakannya secara baik dan benar sesuai dengan kaidah baku.

Namun, merujuk pada pendapat Koentjaraningrat dalam Uswati (2015), kebanyakan orang Indonesia yang berbahasa Indonesia berprinsip “*pokoknya mengerti*”. Umumnya, karena ketidaksengajaan atau ketidaktahuan akan kaidah yang berlaku, sering terjadi penyimpangan atau kekeliruan dalam penulisan maupun pengucapan bahasa Indonesia oleh bangsa Indonesia sendiri. Salah satu penyimpangan yang sering terjadi di antaranya adalah struktur kalimat yang rancu (Uswati, 2015).

Jika bangsa Indonesia yang sehari-harinya menggunakan bahasa Indonesia saja dapat melakukan kekeliruan, terlebih lagi bangsa asing (*second language learner*) yang awam dengan kaidah dan tatabahasa baku Bahasa Indonesia. Munawarah (1996) mencatat tiga jenis kesalahan yang dilakukan oleh pembelajar asing pada saat itu meliputi (1) kesalahan memilih kata untuk mewakili konsep-konsep, (2) kesalahan di bidang ejaan, dan (3) kesalahan tata bahasa yang terdiri atas kesalahan imbuhan, kesalahan aktif-pasif, kesalahan konjungsi dan preposisi, serta kesalahan susunan kalimat, yang mana hal-hal tersebut dapat menjadi sebab rancunya suatu kalimat.

Pusat Pembinaan dan Pengembangan Bahasa (2012) menjabarkan bahwa kalimat rancu adalah kalimat yang kacau atau kalimat yang susunannya tidak sesuai dengan kaidah Bahasa Indonesia yang benar, sehingga informasinya sulit

dipahami. Pada bahasa Inggris, kasus kerancuan ini disebut dengan *grammatical errors*.

Kerancuan dapat terjadi pada kelompok kata dan kalimat. Bentuk kerancuan dapat dilihat dari dua sisi, yaitu: (1) sisi struktural/sintaksis; dan (2) sisi makna atau semantik. Dari segi makna atau gagasan, kerancuan pada kalimat bisa terjadi karena terdapat lebih dari satu gagasan yang tergabung pada satu ungkapan atau kalimat, sedangkan secara sintaksis, kerancuan tersebut timbul karena penggabungan lebih dari satu struktur kalimat, ke dalam satu struktur (Arifin & Hadi, 2001).

Beberapa contoh kalimat rancu yang dapat dikategorikan pada sisi bentuk atau struktur adalah sebagai berikut (Pusat Pembinaan dan Pengembangan Bahasa, 2012; Supriyadi, 2014).

- 1) Terdapat pemilihan afiks (imbuhan) yang salah, atau pembentukan kata berimbuhan (afiksasi) yang salah.

Contohnya :

*Jalan masuk motor Universitas Pendidikan Indonesia akan diperlebarkan.*

Kata '*diperlebarkan*' adalah kata-kata yang membuat struktur kalimat-kalimat tersebut menjadi rancu, karena penggunaan imbuhan yang tidak tepat. Contoh yang benar adalah '*diperlebar*'.

- 2) Kalimat tidak memiliki unsur SPO yang lengkap. Contohnya :

*a. Ibu Murni kepada anaknya.*

*b. Menggaji karyawannya.*

Contoh *a)* tidak memiliki predikat yang menjelaskan subjek Ibu Murni. Kalimat tersebut tidak jelas maksudnya sehingga dikategorikan sebagai kalimat rancu. Sementara contoh *b)* termasuk kalimat rancu karena tidak memiliki subjek. Akan lebih baik jika kalimat *c)* diganti dengan '*Pak Farid harus menggaji karyawannya.*', atau '*Bos menggaji karyawannya.*'

- 3) Terdapat kata atau frasa yang tidak jelas fungsi dan perannya di dalam kalimat.

Contohnya :

*Dalam penelitian ini membahas efektivitas algoritma Naive Bayes.*

Fungsi frasa '*Dalam penelitian ini*' dapat dikatakan sebagai keterangan, namun kemudian diikuti oleh kata kerja aktif '*membahas*' yang jelas berfungsi sebagai predikat, sehingga fungsi frasa tersebut seolah-olah menjadi subjek yang berperan sebagai keterangan juga. Hal ini berdasarkan pada makalah Supriyadi (2014) yang menyebutkan bahwa subjek merupakan unsur pokok pada kalimat yang terletak di samping unsur predikat, dan predikat pada kalimat berperan dalam menjelaskan subjek.

Bahasa yang baik dan benar tentu akan meningkatkan kualitas pembicaraan ataupun tulisan seperti karya ilmiah, skripsi, *essay*, surat atau e-mail, buku dan sejenisnya. Benar atau salahnya suatu kalimat dalam bentuk teks juga berdampak pada kesan pembaca terhadap penulis, atau terhadap tulisan itu sendiri. Karena itu, hanya menulis e-mail, yang notabene sudah cukup sering dilakukan pun, masih perlu melalui beberapa kali pengecekan kalimat (Penniston, 2009). Contoh lebih besarnya, setiap perusahaan penerbit selalu mempekerjakan seorang editor yang bertugas untuk menemukan kesalahan pada teks, sebelum kemudian mencetak teks atau naskah tersebut dalam kualitas yang paling baik (Bušta, Hlaváčková, Jakubiček, & Pala, 2009).

Kerancuan kalimat akan mudah diketahui dan dianalisis dalam bentuk tertulis atau teks (Uswati, 2015). Setelah dilakukan analisis, Nugraha (2000) menuliskan solusi alternatif yang dapat dilakukan, yakni memberikan remidi atau perbaikan. Remidi ini dimaksudkan sebagai proses penyadaran atas berbagai kesalahan yang dilakukan pembelajar (dalam hal ini pengguna bahasa) untuk kemudian dilakukan berbagai upaya penanggulangan agar kesalahan-kesalahan tersebut tidak terjadi lagi (Norissh, 1983 dalam Nugraha, 2000; Suratminto, 1996). Adapun langkah-langkah remidi tersebut adalah: (1) menginformasikan kesalahan kepada yang bersangkutan, (2) melakukan koreksi, kemudian (3) memberikan contoh-contoh yang benar atas kesalahan-kesalahan tersebut, sehingga pengguna bahasa dapat membandingkan sendiri antara tata bahasa yang keliru dengan tata bahasa yang benar (Nugraha, 2000).

Pada masa kini, sebagian besar teks sudah diproduksi menggunakan komputer. Maka, sejalan dengan perkembangan teknologi komputer yang sangat pesat pada era digital ini, tidak heran jika ada tendensi yang kuat untuk juga menggunakan komputer dalam hal memperbaiki kalimat yang salah dalam suatu teks, agar lebih mempersingkat waktu dan memberi banyak kemudahan bagi pengguna bahasa Indonesia, terutama jika teks yang perlu dianalisis sangat panjang dan banyak.

Implementasi kecerdasan buatan sangat mungkin dalam melakukan analisis dan langkah-langkah remedi yang telah dipaparkan pada paragraf sebelumnya. Dengan mengimplementasikan metode *machine learning* dan *Natural Language Processing* (NLP), Penniston (2009) dapat membangun sistem yang mampu memprediksi *grammatical error* pada suatu teks bahasa Inggris dengan akurasi di atas 70%, Andersen (2011) bahkan telah menambahkan metode untuk *replacement* yang dalam hal ini berarti aktivitas menggantikan *grammatical error* dengan susunan yang benar secara otomatis. Begitupun, berbagai macam program atau perangkat lunak yang bertugas sebagai *spelling checker*, *grammar checker*, hingga *style checker* sudah dipublikasikan. Namun keberadaannya masih terbatas (Bušta, Hlaváčková, Jakubíček, & Pala, 2009), terutama yang digunakan khusus untuk Bahasa Indonesia.

Di sisi lain, mengingat fungsi, sejarah dan kedudukan bahasa Indonesia, sudah selayaknya bangsa Indonesia menunjukkan karakter berbahasa Indonesia secara baik dan benar sebagai perwujudan rasa cinta dan bangga berbahasa Indonesia. Penelitian bagaimana tepatnya metode *natural language processing* dapat digunakan untuk mendeteksi kalimat berstruktur rancu secara otomatis, memungkinkan biaya mengoreksi kemunculan kata berstruktur rancu pada media-media tekstual dapat diminimalisir. Selain itu, dengan adanya pendeteksi kalimat rancu pada bahasa Indonesia, dapat berperan dalam bidang penelitian NLP yang lebih besar. Di antaranya adalah penerjemahan otomatis dari bahasa asing ke dalam Bahasa Indonesia, kemudian peringkasan otomatis.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dari penelitian ini adalah:

- a. Bagaimana rancang bangun sistem pendeteksi kalimat berstruktur rancu pada Bahasa Indonesia menggunakan metode pengolahan bahasa alami, *sequential pattern mining* dan klasifikasi teks?
- b. Bagaimana pengujian kelayakan sistem pendeteksi kalimat berstruktur rancu pada Bahasa Indonesia?

## 1.3 Tujuan

Adapun tujuan dari penelitian ini berdasarkan rumusan masalah yang telah disebutkan, adalah sebagai berikut :

- a. Membuat sistem pendeteksi kalimat bahasa Indonesia berstruktur rancu, sebagai alternatif solusi untuk mengetahui kerancuan kalimat dalam suatu teks bahasa Indonesia.
- b. Mengimplementasikan metode pengolahan bahasa alami, *sequential pattern mining* dan klasifikasi teks pada sistem pendeteksi kalimat bahasa Indonesia berstruktur rancu.
- c. Menguji sistem pendeteksi kalimat rancu yang telah dirancang sehingga dapat diukur kelayakannya.

## 1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan sumbangan pemikiran pada aspek teoritis dan praktis sebagai berikut.

- a. Penelitian ini melalui pendekatan dan metode-metode yang digunakan terutama dalam upaya menggali pendekatan-pendekatan baru diharapkan dapat memberikan sumbangan dalam aspek teoritis pada umumnya yang berkaitan dengan Ilmu Komputer, dan khususnya pada bidang pemrosesan bahasa alami (*natural language processing*) pada Bahasa Indonesia, yaitu deteksi kerancuan struktur kalimat dalam Bahasa Indonesia.

- b. Penelitian ini diharapkan dapat memberikan sumbangan pemikiran bagi perusahaan atau komunitas untuk dijadikan pertimbangan dalam membantu pekerjaan editor dalam memperbaiki suatu dokumen dari sisi kerancuan struktur kalimat.
- c. Hasil penelitian diharapkan dapat dijadikan sebagai informasi atau landasan untuk melakukan penelitian selanjutnya mengenai pemrosesan bahasa alami pada Bahasa Indonesia khususnya pada tingkat kalimat, mengingat masih banyak batasan yang ada pada penelitian ini.

### 1.5 Batasan Masalah

Batasan masalah yang ditetapkan pada penelitian ini adalah sebagai berikut.

- a. Bentuk kebahasaan yang diteliti adalah Bahasa Indonesia yang berwujud teks, yaitu artikel pada media massa online, dan karya fiksi.
- b. Sistem yang akan dirancang bekerja pada kesalahan struktur kalimat (sintaksis).
- c. Metode yang digunakan dalam penelitian ini adalah metode *Sequential Pattern Mining* dengan jenis *labeled sequential pattern* (LSP) yang dikombinasikan dengan metode klasifikasi teks.

### 1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini terbagi ke dalam lima bab, sesuai dengan ketentuan yang ditetapkan di Universitas Pendidikan Indonesia. Bab-bab tersebut meliputi Bab I : Pendahuluan; Bab II : Tinjauan Pustaka; Bab III : Metodologi Penelitian; Bab IV : Hasil Penelitian dan Pembahasan; dan Bab V: Kesimpulan dan Saran.

## BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian yang akan dilakukan, serta sistematika penulisan. Latar belakang menceritakan alasan munculnya gagasan penelitian dan pengambilan judul penelitian ini, yakni meliputi pentingnya menggunakan bahasa Indonesia secara baik dan benar tanpa kerancuan, kondisi aktual pengguna bahasa Indonesia—baik

bangsa Indonesia sendiri maupun bangsa asing, serta perlunya implementasi metode *natural language processing* untuk membangun sistem pendeteksi kalimat berstruktur rancu secara otomatis sebagai solusi alternatif analisa kerancuan kalimat dalam teks berbahasa Indonesia.

## **BAB II TINJAUAN PUSTAKA**

Bab ini berisi penjelasan tentang teori-teori dan konsep metode serta algoritma yang digunakan dalam penelitian, meliputi teori dan konsep tentang kalimat pada bahasa Indonesia, kalimat rancu dalam bahasa Indonesia, teori *natural language processing* mulai dari definisi, karakteristik, hingga contoh aplikasi yang ada dan topik riset sebelumnya, cara mengetahui kesalahan kalimat secara sintaksis, *sequential pattern mining* hingga pendekatan dan algoritmanya, *labeled sequence pattern* yang merupakan salah satu jenis representasi *sequential pattern*, serta algoritma *machine learning* yang digunakan dalam topik klasifikasi teks.

## **BAB III METODOLOGI PENELITIAN**

Bab ini berisi penjelasan langkah-langkah yang akan dilakukan dalam penelitian, meliputi desain penelitian, metode penelitian mulai dari metode pengumpulan data teks yang akan diuji, hingga pembangunan sistem pendeteksi kalimat rancu, serta alat dan bahan yang diperlukan untuk melaksanakan penelitian.

## **BAB IV HASIL PENELITIAN DAN PEMBAHASAN**

Berisi penjelasan dari hasil penelitian yang telah dilakukan yaitu proses pengumpulan data penelitian, pengembangan perangkat lunak, dan pengujian.

## **BAB V KESIMPULAN DAN SARAN**

Berisi kesimpulan dan saran yang didapatkan selama pelaksanaan penelitian.