

BAB I

PENDAHULUAN

1.1. Latar Belakang

Bahasa memiliki peranan sentral dalam setiap kehidupan yang dijalani oleh manusia, karena merupakan alat utama untuk berinteraksi dengan sesama. Komunikasi bisa dilakukan melalui bahasa lisan maupun tulisan. Penggunaan bahasa tulisan memungkinkan penyimpanan dan penyaluran pengetahuan dari satu generasi ke generasi berikutnya. Dalam praktek untuk menulis sebuah karya tulis, penting untuk menggunakan kalimat yang efektif. Kalimat yang efektif adalah kalimat yang mengikuti tata bahasa dan kaidah Bahasa Indonesia yang baik dan benar. Hal ini akan berdampak pada kemudahan dan kejelasan dalam menyampaikan ide atau pesan. Seperti yang disampaikan oleh ahli bahasa David Wilkins, “tanpa tata bahasa, sangat sedikit yang bisa disampaikan, dan tanpa kosakata, tidak ada yang bisa disampaikan” [1].

Pentingnya penguasaan bahasa Indonesia juga tercermin dalam kurikulum Pendidikan di Indonesia. Setiap siswa diwajibkan untuk belajar keterampilan berbahasa Indonesia, termasuk mendengarkan, membaca, berbicara, dan menulis. Kemampuan berbahasa seseorang mencerminkan kejernihan dan ketepatan dalam berpikir [2]. Dalam penggunaan bahasa, baik lisan maupun tertulis, penting untuk memahami aturan dan kaidah yang berlaku agar dapat menyusun kalimat yang efektif dalam menyampaikan gagasan, pikiran, atau perasaan. Kamus Besar Bahasa Indonesia (KBBI) merupakan pedoman utama dalam memahami dan mengaplikasikan kaidah bahasa Indonesia [3].

Kemampuan untuk mengenali dan menggunakan pola kalimat dasar sangat penting dalam pengembangan kalimat yang variatif dan sesuai konteks. Dengan memahami dan menerapkan kaidah yang benar, kita dapat menghasilkan kalimat yang jelas, padat, dan mudah dipahami oleh pembaca atau pendengar [4]. Pengetahuan dan penerapan kaidah bahasa merupakan hal yang mendasar dalam

proses penulisan, karena akan memengaruhi kesan dan pemahaman pembaca terhadap teks yang disampaikan.

Dalam penelitian yang dilakukan oleh Titin Nurhayatin dan rekan-rekannya, bersumber dari 25 makalah yang ditulis oleh mahasiswa Prodi Pendidikan Guru Sekolah Dasar (PGSD) masih menemukan kesalahan berbahasa pada penulisan tersebut. Dari analisis tersebut, ditemukan 94 kesalahan berbahasa yang meliputi kesalahan aspek struktur kalimat, kesejajaran, ejaan, diksi, dan kelogisan. Kesalahan terbanyak terjadi dalam penggunaan struktur kalimat yang mendapat hasil kesalahan sebesar 24,7% dari keseluruhan data yang dianalisa [5].

Dari temuan tersebut, jelaslah bahwa pemahaman dan penerapan kaidah bahasa menjadi hal mendasar dalam proses penulisan. Kesalahan-kesalahan berbahasa dapat mengganggu kesan dan pemahaman pembaca terhadap teks yang disampaikan. Oleh karena itu, perlu adanya upaya untuk meningkatkan kemampuan dalam mengenali dan menggunakan pola kalimat dasar, sehingga dapat menghasilkan kalimat yang jelas, padat, dan mudah dipahami.

Dalam konteks ini, teknik yang tepat sangat diperlukan untuk menganalisis pola kalimat. Dalam bidang *Natural Language Processing* (NLP) atau pemrosesan bahasa alami, terdapat berbagai teknik yang dapat digunakan dalam pengenalan kata maupun pola kalimat. Salah satu teknik tersebut adalah LALR (*Look-Ahead Left-Right*) *Parser*, yang merupakan variasi dari LR *Parser*, dan dapat berguna dalam memahami struktur sintaksis suatu teks.

Pada penelitian yang dilakukan oleh Mudafiq Riyan Pratama dan timnya menggunakan *Context Free Grammar* dan LALR *Parser* untuk mendeteksi pola kalimat Bahasa Indonesia. Hasilnya menunjukkan bahwa *Context Free Grammar* memiliki keterbatasan dalam mendeteksi pola kalimat yang lebih kompleks karena unsur pembentuknya tidak selalu tersedia. LALR *Parser* dapat mendeteksi dengan baik semua pola kalimat yang telah didefinisikan sebelumnya, ditunjukkan bahwa tingkat akurasi di atas 75% dari total semua pola kalimat [6]. Namun terkadang suatu kalimat dapat memiliki entri yang didefinisikan ganda, sehingga LALR *Parser* tidak dapat menangani struktur kalimat tersebut dikarenakan terdapat unsur keambiguitas di dalamnya [7].

Part-of-Speech (POS) Tagging merupakan salah satu teknik dalam NLP yang berperan memberikan penanda sintaksis untuk setiap kata dalam sebuah teks berdasarkan konteks dari kata tersebut, sehingga dapat mengenali suatu kalimat tanpa terdapat unsur keambiguitas. Penelitian yang dilakukan oleh Kurnia Rismawati dan rekan-rekannya menggunakan *Hidden Markov Model* untuk POS *Tagging* pada Bahasa Lampung dialek A, dengan hasil akurasi mencapai 95,05% dari keseluruhan 554 kata yang diuji, di mana 529 kata dilabeli dengan benar dan 25 kata salah [8].

Salah satu contoh penelitian yang dilakukan Maharani dan tim menunjukkan potensi penggunaan POS *Tagging* dalam memahami struktur kalimat Bahasa Indonesia tentang Swamedika. Dalam penelitiannya, POS *Tagging* berhasil melabeli 269 dari 300 kata dengan benar dan menghasilkan akurasi rata-rata sebesar 70,37% [9]. Selain itu, penelitian lain yang menggunakan POS *Tagging* dan LALR *Parser* berhasil memeriksa 79,2% kalimat dengan tingkat akurasi 85,1%. Hal ini menunjukkan bahwa kombinasi antara POS *Tagging* dan LALR *Parser* memiliki potensi untuk membantu dalam memahami struktur kalimat Bahasa Indonesia secara lebih efektif [10].

Penelitian ini memanfaatkan POS *Tagging* dengan *Hidden Markov Model* sebagai dasar untuk LALR *Parser* dalam memahami struktur kalimat bahasa Indonesia. Pemahaman terhadap struktur kalimat dasar menjadi faktor penting dalam membangun kalimat yang jelas dan tepat. Pengetahuan dan penerapan teknik ini menjadi bagian penting dalam meningkatkan kemampuan berbahasa dan pemrosesan teks secara lebih baik.

1.2. Rumusan Masalah

1.2.1. Permasalahan

Dengan seiring dengan perkembangan zaman, kalimat yang digunakan dalam berbahasa menjadi semakin kompleks. Diperlukan metode untuk melakukan pemeriksaan kalimat yang kompleks sesuai dengan kaidah atau aturan bahasa yang berlaku. LALR (*Look-Ahead Left-Right*) *Parser* tidak dapat melakukan pemeriksaan

struktur kalimat berbahasa Indonesia dengan efektif dikarenakan terdapat entri ganda yang akan membuat hasil parsing menjadi ambigu.

1.2.2. Metode Usulan

Dengan menggunakan *Hidden Markov Model* dari *POS Tagging*, tiap kata dapat diseleksi dengan baik sesuai dengan jenis kata tersebut dan aturan bahasa yang berlaku. Kombinasi *POS Tagging* dan *LALR Parser* dapat dijadikan sebagai eksperimen untuk menguji keefektifitas dalam memahami pola kalimat berbahasa Indonesia agar mendapatkan nilai akurasi pemeriksaan yang baik.

1.2.3. Pertanyaan Penelitian

Bagaimana hasil akurasi dari proses kombinasi *POS Tagging* dan *LALR Parser* untuk memahami dan meningkatkan kemampuan dalam pemeriksaan struktur kalimat dalam konteks bahasa Indonesia?

1.3. Batasan Masalah

Penelitian ini memiliki beberapa batasan untuk memastikan fokus dan ketetapan dalam mencapai tujuan penelitian. Berikut adalah beberapa batasan yang diterapkan:

1. Berfokus pada penerapan *POS Tagging Hidden Markov Model* dan *LARL Parser*.
2. Data yang digunakan adalah kalimat dari teks berita.
3. Evaluasi akan dilakukan berdasarkan kemampuan dalam mengenali kalimat.
4. Panjang kalimat yang diproses berdasarkan data yang digunakan.
5. Dalam sistem hanya terdapat 2 kategori sebagai hasil yaitu benar atau salah.

1.4. Tujuan dan Manfaat

1.4.1. Tujuan

Penelitian ini bertujuan untuk menguji dan mengetahui seberapa besar akurasi yang didapat dari kombinasi *POS Tagging* dan *LALR Parser* dalam memahami dan mengembangkan struktur kalimat bahasa Indonesia pada data teks berita.

1.4.2. Manfaat

Penelitian ini diharapkan membawa manfaat bagi kemajuan teknologi Pemrosesan Bahasa Alami, terutama dalam bahasa Indonesia. Hasil dari penelitian ini dapat menjadi landasan untuk meningkatkan kemampuan berbahasa dan pemrosesan teks pada berbagai aplikasi seperti pengecekan tata bahasa otomatis, pengenalan suara, dan mesin penerjemah.