

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Teknologi selalu berkembang dan sedikit demi sedikit akan mengubah perilaku manusia dalam kesehariannya. Menurut Nazwa Salsabila dan Muhammad(2023), Perkembangan teknologi informasi memainkan peran sentral dalam perubahan masyarakat modern [5]. Salah satu perubahan yang dipengaruhi oleh perkembangan teknologi adalah dalam bidang komunikasi, seperti cara manusia berkomunikasi yang sebelumnya hanya dapat dilakukan dengan bertemu secara langsung atau secara tidak langsung menggunakan surat, tetapi saat ini manusia dapat berinteraksi dengan cepat dengan siapa saja dan dimana saja menggunakan teknologi. Teknologi Komunikasi dan Informasi adalah aplikasi pengetahuan dan keterampilan yang digunakan manusia dalam mengalirkan informasi atau pesan dengan tujuan untuk membantu menyelesaikan permasalahan manusia (aktivitas sosial) agar tercapai tujuan komunikasi [6]. Tidak hanya kehidupan manusia yang dipengaruhi oleh perkembangan teknologi, tetapi perkembangan teknologi juga dapat dipengaruhi oleh kehidupan manusia. Salah satu teknologi yang dikembangkan adalah pengembangan teknologi mobil otonom menggunakan kecerdasan buatan. Mobil otonom dapat digunakan untuk membantu kegiatan manusia dan mengurangi interaksi antar manusia secara langsung saat pandemi berlangsung seperti transportasi manusia dan barang [7]. Kecerdasan buatan tidak hanya digunakan untuk membuat mobil otonom, tetapi kecerdasan buatan juga dapat digunakan untuk membantu manusia dalam bidang bahasa.

Pada topik kecerdasan buatan yang difokuskan dalam bidang bahasa dan digunakan untuk mengolah bahasa dapat disebut dengan *Natural Language Processing* (NLP). NLP digunakan untuk membantu mesin memahami bahasa alami manusia (*Natural Language*) [8]. Saat ini, NLP sudah diaplikasikan ke dalam berbagai macam alat yang dapat membantu pekerjaan manusia. *Automated Customer Service* [9], aplikasi pendeteksi kesalahan bahasa [10, 11, 12, 13], aplikasi pengenalan suara [14], dan aplikasi pendeteksi spam [15] adalah beberapa bentuk pengaplikasian NLP yang dapat membantu kegiatan manusia. NLP dimulai sejak tahun 1950 dan terus mengalami perkembangan hingga saat ini. Ada berbagai macam metode dan algoritma dibuat untuk meningkatkan performa NLP. Beberapa

algoritma dan metode NLP yang sudah dibuat dan diaplikasikan adalah *Support Vector Machine* (SVM), *Hidden Markov Model* (HMM), *Conditional Random Field* (CRF), dan *N-gram* [16].

Dengan tujuan melanjutkan pengembangan penelitian mengenai NLP, aplikasi U-Tapis dibuat sebagai aplikasi yang dapat membantu jurnalis dan editor/penyunting untuk mendeteksi berbagai macam kesalahan secara tepat dan cepat dalam pekerjaannya [17]. U-Tapis merupakan sebuah aplikasi yang dapat digunakan untuk menjadi alat penyaring kesalahan bahasa dengan tujuan untuk menjadi sarana penunjang bagi calon jurnalis dan para jurnalis untuk mempermudah pengecekan kesalahan penulisan Bahasa Indonesia pada artikel berita [18]. U-Tapis melakukan analisis berdasarkan kumpulan data tekstual dalam jumlah besar. Terdapat berbagai analisis yang dilakukan, diantaranya adalah analisis statistik, bahasa komputasi, dan kecerdasan buatan [18]. Hasil analisis yang dilakukan pada aplikasi U-Tapis dapat digunakan untuk mengidentifikasi dan memperbaiki kesalahan tata bahasa, ejaan, dan struktur kalimat dalam artikel berita. Hal tersebut dapat meningkatkan kualitas artikel yang dihasilkan jurnalis dan meringankan pekerjaan editor/penyunting. Penelitian telah dilakukan untuk menilai tingkat efektivitas dari aplikasi U-Tapis dengan melakukan pengujian terhadap mahasiswa dari suatu universitas swasta untuk membuat teks berita dan artikel opini sebelum dan sesudah menggunakan aplikasi U-Tapis, kemudian dilakukan penilaian terhadap teks tersebut dan dilakukan perbandingan nilai teks sebelum dan sesudah menggunakan aplikasi U-Tapis. Dalam penelitian tersebut ditemukan peningkatan nilai yang signifikan sehingga dapat disimpulkan aplikasi U-Tapis efektif untuk digunakan sebagai alat pendeteksi kesalahan bahasa dan layak untuk dikembangkan lebih lanjut [17].

Salah satu faktor yang harus diperhatikan dalam penggunaan bahasa yang benar adalah penggunaan kata konjungsi. Konjungsi adalah kata atau ungkapan penghubung antarkata, antarfrasa, antarklausa, dan antarkalimat [19]. Berdasarkan definisi dari Kamus Besar Bahasa Indonesia (KBBI), maka konjungsi digunakan sebagai penghubung antar kalimat yang masih berhubungan dengan kalimat sebelumnya. Hal ini membuat konjungsi berguna untuk memperjelas kalimat yang disampaikan karena dapat mengurangi kata-kata yang memiliki arti sama dengan kalimat sebelumnya. Tetapi terdapat beberapa aturan yang harus dipatuhi dalam penggunaan kata konjungsi ini agar kalimat yang menggunakan konjungsi tersebut dapat tetap menjadi kalimat efektif dan pesan yang ingin disampaikan bisa tersampaikan dengan baik. Meskipun begitu, masih dapat

ditemukan kesalahan dalam penggunaan kata konjungsi tersebut. Dalam penulisan artikel berita, kata konjungsi cukup banyak digunakan karena sebuah berita yang baik harus memperhatikan hubungan antarkalimat untuk memelihara keterkaitan dan keruntutan antarkalimat [20]. Jika sebuah artikel ingin dipahami oleh pembaca dengan baik, ide atau gagasan harus dihubungkan dengan baik antarkalimat karena kesalahan dalam penggunaan kata konjungsi dapat merusak makna pada suatu wacana [21]. Banyaknya kata konjungsi yang digunakan dapat meningkatkan potensi terjadi kesalahan dalam penggunaannya. Potensi kesalahan tersebut juga dapat meningkat dengan jumlah artikel yang harus dibuat oleh seseorang dalam waktu yang berdekatan. Jurnalis untuk media daring TribunNews.com diharuskan untuk memproduksi setidaknya 20 artikel setiap harinya [18]. Proses penyuntingan artikel dapat memakan waktu dan tenaga, terutama ketika artikel yang harus melalui proses penyuntingan mencapai 20 artikel setiap harinya. Proses tersebut dapat dipermudah menggunakan aplikasi U-Tapis sekaligus mengurangi potensi adanya kesalahan yang terlewati.

Dalam penelitian ini, sistem deteksi kesalahan penggunaan kata konjungsi dibuat dengan menggunakan *Semantic Analysis* dan saran perbaikan dibuat menggunakan model BERT. Deteksi kata konjungsi sebelumnya dilakukan menggunakan algoritma *cosine similarity* oleh Olwen untuk deteksi kesalahan eja kata konjungsi dengan tingkat akurasi 92.2%, tetapi algoritma tersebut tidak dapat menghasilkan saran perbaikan berdasarkan hasil deteksi. Metode *Semantic Analysis* digunakan untuk melakukan deteksi kesalahan penggunaan kata konjungsi karena metode *Semantic Analysis* dapat digunakan untuk menghindari ambiguitas pada kalimat [22]. Sedangkan model BERT sudah pernah digunakan untuk menghasilkan saran perbaikan untuk kesalahan peluluhan kata oleh Kevin dengan tingkat akurasi 88.89%, 85.23%, and 84.68% dalam pengujian terhadap 100, 150, dan 200 artikel [23]. Model BERT juga pernah digunakan untuk membuat *POS tagger* untuk Bahasa Indonesia dan berhasil memberikan label yang benar untuk kata ambigu dengan pencapaian sebesar 96/100 [24]. Penerapan *Semantic Analysis* dilakukan dengan membuat daftar aturan penggunaan kata konjungsi dan daftar kata konjungsi beserta aturan yang perlu dipenuhi oleh setiap kata tersebut. Setelah deteksi dilakukan, maka setiap kalimat yang mengandung kata konjungsi dan terdeteksi kesalahan dalam penggunaannya akan diproses kembali menggunakan model BERT untuk menghasilkan saran perbaikan dalam bentuk prediksi kata berikutnya berdasarkan konteks dari kalimat.

Terdapat beberapa saran juga dalam penelitian deteksi kata konjungsi yang

dilakukan oleh Olwen untuk meningkatkan deteksi kesalahan kata konjungsi. Saran yang diberikan adalah meningkatkan tahap iterasi data agar sistem dapat mendeteksi konjungsi yang terdiri lebih dari satu kata serta mengembangkan algoritma klasifikasi agar sistem dapat mengenali makna kata konjungsi sehingga sistem deteksi tidak hanya berfokus pada sifat konjungsi antara dan konjungsi awalan [25].

## 1.2 Rumusan Masalah

1. Bagaimana *Semantic Analysis* dapat digunakan untuk melakukan deteksi penggunaan kata konjungsi dan model BERT untuk memberikan saran perbaikan penggunaan kata konjungsi pada artikel berita?
2. Berapa tingkat akurasi *Semantic Analysis* dalam mendeteksi kesalahan penggunaan kata konjungsi pada artikel berita menggunakan *Confusion Matrix*, *Precision Recall* dan *F1-Score* dan model BERT untuk memberikan saran perbaikan penggunaan kata konjungsi dalam artikel berita?

## 1.3 Batasan Permasalahan

1. Hasil akhir pengembangan berupa modul sebagai salah satu fitur dari situs U-Tapis.
2. Pembuatan model menggunakan data yang disediakan oleh tim dari Program Studi Jurnalistik UMN.
3. Nilai *input* berupa teks artikel berita yang tidak memiliki kesalahan dalam Bahasa Indonesia selain kesalahan penggunaan kata konjungsi.
4. Hasil *output* berupa teks berita yang setiap penggunaan kata konjungsi yang benar diberikan warna hijau dan setiap penggunaan kata konjungsi yang salah berupa *dropdown* dengan daftar pilihan saran perbaikan.

## 1.4 Tujuan Penelitian

1. Membuat sebuah model *Semantic Analysis* yang dapat mendeteksi kesalahan penggunaan kata konjungsi dan model BERT untuk membuat saran perbaikan penggunaan kata konjungsi pada artikel berita.

2. Mengukur tingkat akurasi model yang dapat mendeteksi kesalahan penggunaan kata konjungsi dalam artikel berita menggunakan *Confusion Matrix*, *Precision Recall* dan *F1-Score* dan model BERT .

### 1.5 Manfaat Penelitian

1. Bagi peneliti
  - (a) Dapat mengimplementasikan teori selama perkuliahan, terutama tentang *Natural Language Processing* dengan menggunakan metode *Semantic Analysis* serta model BERT.
2. Bagi pihak lain
  - (a) Dapat membantu jurnalis untuk membuat artikel berita dengan penggunaan kata konjungsi yang benar.
  - (b) Dapat memberikan informasi mengenai deteksi penggunaan kata konjungsi menggunakan metode *Semantic Analysis* serta menggunakan model BERT untuk membuat saran perbaikan kata konjungsi untuk penelitian selanjutnya sehingga lebih efektif.

### 1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN  
Bab 1 berisi penjelasan latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan
- Bab 2 LANDASAN TEORI  
Bab 2 berisi penjelasan teori yang digunakan sebagai dasar pada laporan skripsi ini. Teori yang digunakan adalah kata konjungsi, *Semantic Analysis*, model BERT, dan *Confusion Matrix*.
- Bab 3 METODOLOGI PENELITIAN  
Bab 3 berisi penjelasan mengenai metode yang digunakan dalam

penelitian ini. Langkah-langkah metode yang dilakukan dalam penelitian ini adalah studi literatur, pengumpulan dan praproses data artikel, pembuatan *rule* penggunaan kata konjungsi, perancangan algoritma deteksi kesalahan penggunaan kata konjungsi, perancangan algoritma rekomendasi menggunakan model BERT, perancangan API, dan pengujian dan evaluasi terhadap performa algoritma deteksi kesalahan serta performa API.

- Bab 4 HASIL DAN DISKUSI

Bab 4 berisi penjelasan implementasi metode penelitian yang ada pada bab 3 serta hasil implementasi dari algoritma deteksi kesalahan penggunaan kata konjungsi menggunakan *semantic analysis* dan model BERT pada untuk aplikasi U-Tapis.

- Bab 5 KESIMPULAN DAN SARAN

Bab 5 berisi kesimpulan yang didapatkan dari penelitian ini dan saran untuk penelitian serupa di masa depan.

