

BAB I

PENDAHULUAN

1.1 Latar Belakang

Test of English as a Foreign Language (TOEFL) merupakan landasan dalam mengevaluasi kemahiran akademik bahasa Inggris bagi non-penutur asli, dan berfungsi sebagai pintu gerbang untuk masuk ke universitas dan institusi di seluruh dunia (Alderson, 2000). Tes standar ini terdiri dari empat bagian, masing-masing dirancang untuk menilai keterampilan bahasa tertentu yang penting untuk keberhasilan akademik. Di antaranya, bagian *Error Identification* memainkan peran penting dalam mengevaluasi penguasaan tata bahasa Inggris seorang kandidat (Bachman & Palmer, 1996). Bagian ini menyajikan kalimat kepada peserta tes, beberapa di antaranya mengandung kesalahan tata bahasa, sehingga mengharuskan mereka untuk mengidentifikasi dan memperbaiki ketidakakuratan ini. Secara tradisional, analisis kinerja siswa pada bagian ini sangat bergantung pada evaluasi manual oleh penguji terlatih (Chodorow & Leacock, 2000). Namun, pendekatan ini mempunyai keterbatasan yang melekat. Proses ini pada dasarnya memakan waktu dan mengharuskan penguji menganalisis setiap respons dengan cermat, sehingga dapat menyebabkan keterlambatan dalam memberikan umpan balik kepada peserta tes. Selain itu, analisis manual pada dasarnya bersifat subjektif, rentan terhadap variasi interpretasi pemeriksa dan potensi bias, sehingga berdampak pada keandalan dan konsistensi penilaian (Jacobs & others, 1981). Sifat keterlibatan manusia menimbulkan risiko kesalahan, terutama ketika menangani data pengujian dalam jumlah besar. Kelemahan ini menggarisbawahi perlunya pendekatan yang lebih efisien, obyektif, dan otomatis untuk mengevaluasi keterampilan tata bahasa dalam tes standar seperti TOEFL.

Penilaian yang akurat terhadap kemahiran berbahasa merupakan faktor penting dalam keberhasilan akademik siswa internasional, yang secara langsung berdampak pada kemampuan mereka untuk berkembang di lingkungan pendidikan tinggi (Andrade, 2006). Penelitian secara konsisten menunjukkan korelasi yang kuat antara kemahiran berbahasa dan kinerja akademik di berbagai disiplin ilmu (misalnya, (Cummins, 2000); Shohamy, 2020). Sebagaimana disoroti oleh Ellis

(1997) "Bahasa bukan sekadar alat komunikasi tetapi media yang melaluinya pembelajaran berlangsung". Ketika penilaian bahasa gagal untuk secara akurat menangkap kemampuan sebenarnya seseorang, karena faktor-faktor seperti varians konstruk yang tidak relevan atau bias budaya, maka konsekuensinya dapat merugikan. Penilaian yang tidak akurat atau bias dapat menyebabkan siswa salah penempatan dalam program akademik yang tidak sesuai, membatasi akses mereka terhadap tugas-tugas yang menantang dan menghambat pertumbuhan akademik mereka (Chiswick & Miller, 2007). Lebih jauh lagi, penilaian semacam ini dapat melanggengkan kesenjangan dengan cara merugikan siswa yang berasal dari latar belakang linguistik tertentu secara tidak adil, sehingga tidak memberi mereka kesempatan yang sama untuk menunjukkan pengetahuan dan keterampilan mereka (Kunnan, 2000). Hal ini menggarisbawahi pentingnya etika untuk mengembangkan dan menerapkan penilaian bahasa yang dapat diandalkan dan valid, memastikan keadilan dan kesetaraan dalam kesempatan pendidikan bagi semua siswa, terlepas dari latar belakang bahasa mereka.

Analisis manual terhadap pertanyaan Identifikasi Kesalahan, meskipun merupakan praktik yang sudah berlangsung lama, menghadirkan serangkaian tantangan unik yang menghambat efisiensi dan objektivitas penilaian bahasa. Salah satu permasalahan yang menonjol terletak pada subjektivitas yang melekat dalam penilaian manusia. Variabilitas antar penilai, sebuah fenomena yang terdokumentasi dengan baik dalam pengujian bahasa (misalnya, Weigle, 2002), muncul ketika banyak penguji berbeda dalam interpretasi aturan tata bahasa dan penerapannya, sehingga menyebabkan pola penilaian yang tidak konsisten. Sebagaimana disoroti dalam Manual Teknis TOEFL® "Meskipun telah menjalani pelatihan yang ketat, penilaian manusia terhadap kemahiran bahasa mungkin berbeda". Variabilitas ini tidak hanya berdampak pada reliabilitas nilai tes tetapi juga menimbulkan kekhawatiran mengenai keadilan, karena kinerja siswa mungkin terlalu dipengaruhi oleh interpretasi subjektif dari penguji tertentu. Selain itu, keahlian manusia, meskipun berharga, pada dasarnya dibatasi oleh kendala kognitif dan kerentanan terhadap kelelahan, terutama ketika ditugaskan menganalisis data pengujian dalam jumlah besar. Keterbatasan ini dapat menyebabkan penurunan akurasi dan peningkatan waktu pemrosesan, yang selanjutnya menghambat

efisiensi analisis manual. Meskipun sistem penilaian berbantuan komputer telah muncul sebagai alternatif yang potensial, sistem tersebut sering kali mengalami kesulitan dengan struktur tata bahasa yang berbeda dan mungkin gagal untuk menangkap seluruh kompleksitas kemahiran berbahasa (Chapelle & Douglas, 2006)

Kemampuan linguistik yang melekat pada Natural Language Processing (NLP) menjadikannya alat yang berpotensi transformatif untuk menganalisis dan memahami nuansa tugas penilaian bahasa seperti bagian Identifikasi Kesalahan TOEFL (Yadav dkk., 2020). Dengan mengekstraksi serangkaian fitur linguistik dari setiap pertanyaan, NLP dapat memberikan wawasan berharga ke dalam struktur tata bahasa tertentu dan jenis kesalahan yang sedang dinilai. Misalnya, teknik NLP dapat mengidentifikasi bagian-bagian ucapan, mengurai struktur kalimat, dan mendeteksi anomali dalam pola sintaksis, secara efektif menunjukkan dengan tepat elemen tata bahasa yang ditargetkan oleh setiap pertanyaan (Tetreault dkk., 2013). Analisis terperinci ini dapat membantu pengembang tes merancang penilaian yang lebih fokus dan efektif, memastikan bahwa item tes secara akurat mencerminkan konstruksi tata bahasa yang diinginkan. Investigasi di masa depan perlu mengevaluasi secara ketat efektivitas penilaian berbasis NLP dalam hal korelasinya dengan penilaian manusia dan dampaknya terhadap reliabilitas dan validitas tes (Yannakoudakis dkk., 2011) Selain itu, memastikan keadilan di berbagai latar belakang bahasa adalah hal yang sangat penting, sehingga memerlukan penelitian terhadap potensi bias yang tertanam dalam model NLP dan pengembangan strategi mitigasi (Blodgett dkk., 2016). Terakhir, kelayakan praktis penerapan sistem penilaian berbasis NLP dalam alur kerja penilaian yang ada memerlukan pertimbangan yang cermat.

Penilaian bahasa yang akurat merupakan landasan pemerataan pendidikan, khususnya bagi penutur bahasa Inggris non-pribumi yang mencari peluang akademik di seluruh dunia (Shohamy, 2020). Kemampuan untuk mengevaluasi kemahiran bahasa secara tepat sangat penting untuk membuat keputusan yang tepat mengenai penerimaan, penempatan, dan dukungan pengajaran. Namun, metode penilaian bahasa tradisional, yang seringkali bergantung pada analisis manual

terhadap respon siswa, pada dasarnya dibatasi oleh subjektivitas, keterbatasan waktu, dan potensi kesalahan manusia (Bachman & Palmer, 1996). Keterbatasan ini menggarisbawahi kebutuhan mendesak akan pendekatan yang lebih efisien, obyektif, dan berteknologi maju. *Natural Language Processing* (NLP) menawarkan potensi transformatif untuk merevolusi analisis tugas penilaian bahasa seperti bagian Identifikasi Kesalahan TOEFL. Kapasitas NLP untuk membedah struktur linguistik, mengidentifikasi pola tata bahasa, dan mendeteksi kesalahan dengan akurasi yang meningkat (C. Manning & Schutze, 1999) membuka jalan baru untuk penilaian otomatis, penyediaan umpan balik terperinci, dan analisis item tes yang berbeda. Namun, penerapan NLP dalam penilaian berisiko tinggi bukannya tanpa tantangan. Memastikan keadilan algoritmik, memitigasi potensi bias yang tertanam dalam data pelatihan, dan menjaga transparansi dalam proses penilaian otomatis merupakan pertimbangan etika penting yang memerlukan pemeriksaan yang cermat (Hutchinson & Mitchell, 2018).

Penelitian terkini telah menunjukkan bahwa aplikasi *Natural Language Processing* (NLP) dalam penilaian bahasa otomatis memiliki potensi yang signifikan. (Biswas dkk., 2014) mengusulkan metode klasifikasi pertanyaan yang ringkas dan efektif untuk sistem *Question Answering* dengan mengkategorikan pertanyaan ke dalam tiga kategori luas berdasarkan struktur sintaksis dan jenis jawaban yang diharapkan. Lee (2009) menunjukkan bahwa *Statistical Machine Translation* (SMT) dapat digunakan untuk mengoreksi kesalahan gramatikal pada teks penutur asing, memperlakukan teks yang tidak gramatikal sebagai bahasa sumber dan teks yang gramatikal sebagai bahasa target. Burstein dkk. (2013) menjabarkan perkembangan terkini dalam identifikasi dan koreksi kesalahan gramatikal secara otomatis, dengan fokus pada penerapan pembelajaran mesin untuk meningkatkan akurasi dan efektivitas dalam mendukung pelajar bahasa. He dkk. (2009) mengembangkan sistem penilaian ringkasan otomatis yang menggabungkan *Latent Semantic Analysis* (LSA) dan *n-gram co-occurrence*, menghasilkan akurasi yang tinggi dan kemampuan umpan balik adaptif. Leacock dkk., 2014 memberikan tinjauan komprehensif dan terkini tentang bidang Deteksi Kesalahan Gramatikal Otomatis (AGED), yang mengeksplorasi teknik-teknik mutakhir, jenis kesalahan spesifik, metrik evaluasi, dan aplikasi praktis untuk

mendukung pelajar bahasa. Tetreault dkk., 2013 melaporkan hasil dari *Native Language Identification Shared Task* pertama, menunjukkan kemajuan signifikan dalam identifikasi bahasa asli otomatis dan mengidentifikasi arahan untuk penelitian di masa depan. Studi-studi ini secara kolektif menyoroti potensi transformatif NLP dalam mengotomatiskan dan meningkatkan berbagai aspek penilaian bahasa.

Tujuan dari penelitian ini adalah untuk mengembangkan sistem berbasis *Natural Language Processing* (NLP) yang mampu mengekstraksi fitur kompetensi gramatikal yang diuji dalam soal *Error Identification* pada tes TOEFL. Sistem ini diharapkan dapat mengotomatiskan proses analisis soal, mengidentifikasi pola kesalahan gramatikal, dan meningkatkan efisiensi dalam pengembangan dan evaluasi soal TOEFL. Dengan demikian, sistem ini berpotensi untuk membantu pengajar dan pengembang tes dalam menciptakan materi pembelajaran dan penilaian yang lebih efektif dan terarah untuk meningkatkan kemampuan tata bahasa Inggris penutur asing.

1.2 Rumusan Masalah

Berikut adalah rumusan masalah yang menjadi fokus pembahasan dalam penelitian ini:

1. Bagaimana teknik *Natural Language Processing* (NLP) digunakan untuk mengekstrak fitur dari pertanyaan *Error Identification* TOEFL secara otomatis?
2. Bagaimana hasil dari ekstraksi fitur kompetensi menggunakan NLP pada pertanyaan *Error Identification* TOEFL?
3. Bagaimana tingkat efektivitas teknik NLP untuk mengidentifikasi fitur dan kompetensi dalam soal *Error Identification* TOEFL?

1.3 Tujuan Penelitian

Dari rumusan masalah yang telah dijelaskan sebelumnya, terdapat beberapa tujuan dari penelitian ini adalah sebagai berikut:

1. Merancang model komputasi *Natural Language Processing* untuk ekstraksi fitur kompetensi pada soal *Error Identification* TOEFL.

2. Mengimplementasikan model komputasi *Natural Language Processing* pada proses ekstraksi fitur kompetensi pada soal *Error Identification* TOEFL.
3. Melakukan eksperimen dan analisa terkait keakuratan hasil dalam melakukan ekstraksi fitur kompetensi pada soal *Error Identification* TOEFL menggunakan *Natural Language Processing*.

1.4 Manfaat Penelitian

Adapun manfaat penelitian dalam penelitian ini adalah:

1. Membuat model komputasi yang dapat mempermudah peneliti pada bidang linguistik untuk menganalisa soal *Error Identification* TOEFL.
2. Melakukan model program yang dapat dikembangkan pada penelitian selanjutnya.
3. Memperoleh data hasil ekstraksi fitur kompetensi soal-soal *Error Identification* TOEFL yang dapat digunakan untuk *machine learning*, sistem rekomendasi, dll.

1.5 Batasan Masalah

Untuk menjaga fokus dan ketepatan penelitian, penelitian ini perlu dibatasi. Mengingat luasnya studi kasus yang diteliti, berikut adalah beberapa batasan yang diterapkan:

1. Penelitian ini hanya meneliti fitur dan kompetensi untuk bagian *Error Identification* pada TOEFL ITP.
2. Jumlah kata *underline* pada soal berjumlah 4 kata.
3. Penelitian ini hanya mengidentifikasi 13 kompetensi yang telah ditentukan sebelumnya, dengan fokus pada aspek gramatikal.
4. Analisis tidak mencakup aspek makna kata atau semantik dalam soal.

1.6 Sistematika Penulisan

Pada dokumen skripsi ini terdapat sistematika yang digunakan sebagai pedoman penulisan. Sistematika tersebut dibagi menjadi lima bab, yaitu:

1. BAB I PENDAHULUAN

Bab ini berisi mengenai penjelasan latar belakang permasalahan penelitian ini. Pada Bab ini juga dijelaskan mengenai rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dalam penelitian, dan sistematika penulisan skripsi.

2. BAB II KAJIAN PUSTAKA

Bab ini berisi mengenai penjelasan teori-teori yang digunakan dalam menyelesaikan permasalahan penelitian ini. Adapun teori yang digunakan yaitu teori Astronomi, *Data Stream*, Deteksi Anomali dan Algoritma PEWMA.

3. BAB III METODE PENELITIAN

Bab ini berisi metode dan langkah-langkah yang digunakan dalam penelitian. Pada bab ini juga menjelaskan mengenai instrumen yang digunakan, tahapan pengumpulan data yang dilakukan, hingga langkah-langkah analisis data yang dijalankan.

4. BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi penjelasan mengenai proses dari setiap tahap penelitian. proses tersebut meliputi pengumpulan data, rancangan model komputasi, pengembangan perangkat lunak, rancangan skenario eksperimen dan pembahasan hasil penelitian.

5. BAB V KESIMPULAN DAN SARAN

Bab ini berisi mengenai kesimpulan yang dibuat berdasarkan hasil penelitian. Kemudian terdapat beberapa saran yang dapat digunakan di dalam penelitian selanjutnya.