

BAB I

PENDAHULUAN

1.1 Latar Belakang

Evaluasi dalam pembelajaran ialah saksi terhadap pertumbuhan dan kemajuan siswa kearah tujuan-tujuan atau nilai-nilai yang telah ditetapkan di dalam kurikulum (Wrightstone, 1956). Evaluasi merupakan salah komponen penting dan tahap yang harus ditempuh oleh guru untuk mengetahui keefektifan pembelajaran. Hasil yang diperoleh dapat dijadikan balikan (*feed-back*) bagi guru dalam memperbaiki dan menyempurnakan program dan kegiatan pembelajaran (Arifin, 2012). Sementara itu, Anthony J.Nitko (1996) menjelaskan penilaian adalah suatu proses untuk memperoleh informasi yang digunakan untuk membuat keputusan tentang peserta didik, kurikulum, program, dan kebijakan pendidikan.

Tes adalah alat atau prosedur khusus atau teknik yang digunakan untuk memperoleh tanggapan dari siswa untuk mendapatkan informasi yang memberikan dasar untuk membuat penilaian atau evaluasi mengenai beberapa karakteristik seperti kebugaran, keterampilan, pengetahuan dan nilai (Barrow dan Gee, 2000). Tes tertulis terdiri dari dua jenis yaitu tes objektif dan tes Esai. Tes objektif sering juga disebut tes dikotomi (*dichotomously scored item*) karena jawabannya antara benar atau salah dan skornya antara 1 atau 0. Disebut tes objektif karena penilaiannya objektif. Kelebihan dari tes ini adalah (1) jawaban bersifat mutlak, karena itu penilaian dapat dilakukan secara objektif (2) pengoreksian dapat dilakukan oleh siapa saja, sekalipun tidak mengetahui dan menguasai materinya (2) pemberian skor dapat dilakukan dengan mudah dan cepat. Siapapun yang mengoreksi jawaban tes objektif hasilnya akan sama karena kunci jawabannya sudah jelas dan pasti. Para guru banyak berpendapat bahwa bentuk uraian memang banyak digunakan karena membuat soalnya relatif lebih mudah, tetapi sulit melakukan penilaian, lebih bersifat subjektif, dan tidak adil (Arifin, 2012). Esai dianggap oleh banyak ahli sebagai metode atau alat yang paling tepat untuk menilai hasil dari kegiatan belajar yang kompleks, karena

penulisan esai akan melibatkan kemampuan siswa untuk mengingat, mengorganisasikan, mengekspresikan dan

mengintegrasikan gagasan yang dimiliki oleh siswa tersebut (Valenti, dkk, 2003). Kebebasan dalam menjawab diizinkan oleh pertanyaan esai sehingga jawaban bisa sangat bervariasi. Siswa mungkin diminta untuk memberikan tanggapan singkat dan tepat, atau mereka mungkin diberi kebebasan besar dalam menentukan bentuk dan ruang lingkup jawaban mereka. Sehingga dalam penilaian esai ini dapat memakan banyak waktu, subjektif, dan cenderung tidak dapat diandalkan (Gronlund, 1998).

Banyak peneliti mengklaim bahwa sifat subjektif dari penilaian esai menimbulkan perbedaan nilai yang diberikan oleh orang yang berbeda, hal ini dirasakan oleh siswa sebagai sumber ketidakadilan. Penggunaan teknologi informasi pun telah banyak mendukung penerapan penilaian bentuk tes objektif tersebut seperti penerapannya dalam *e-learning*. Berdasarkan dari hal tersebut, sistem penilaian esai otomatis dikembangkan. Sistem ini dikembangkan untuk membantu pengajar dalam melakukan penilaian terhadap tes dengan bentuk esai. Penggunaan teknologi informasi tersebut memungkinkan seorang pengajar memangkas waktu dalam proses koreksi terhadap tes bentuk esai, mengurangi tenaga manusia, dan meningkatkan objektivitas dalam melakukan pengkoreksian. Selain itu dengan menggunakan sistem penilaian esai ini hasil penilaiannya dapat menjadi tolak ukur kemampuan siswa dalam memahami pelajaran, dan guru pun dapat mengevaluasi kegiatan pembelajarannya berdasarkan hasil penilaian yang didapat.

Sampai saat ini telah banyak sistem penilaian esai otomatis yang dikembangkan dengan berbagai metode. Perkembangan sistem penilaian esai diawali dengan *Project Essay Grading* (PEG) adalah salah satu pelopor dan bertahan cukup lama pada penilaian esai otomatis dengan menggunakan metode statistik *linguistic feature* (Hearst, 2000; Page, 1994). Pada tahun 2008 telah dilakukan penelitian terhadap penilaian jawaban esai dengan menggunakan *Term Frequency* dan *Inverse Document Frequency* (TF-IDF) dan menghitung kemiripan menggunakan *cosine* dalam algoritma K-Nearest Neighbor (KNN) dengan tingkat akurasi 76% (Bin, dkk, 2008). Sedangkan pada tahun 2010 telah dilakukan penelitian terhadap penilaian jawaban esai pada bahasa China dengan menggunakan beberapa kombinasi Vector Space Model

(VSM), yaitu Latent Semantic-based Vector Space Model (LS-VSM), Sequence Latent Semantic-based Vector Space Model (SLS-VSM), Word-based Vector Space Model (W-VSM), dan Weight Adapted Word-based Vector Space Model (WAW-VSM) (Peng, dkk, 2010). VSM merupakan suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dengan suatu query.

Pada tahun 2012 dilakukan penelitian yang membandingkan beberapa Sistem Penilaian Esai Otomatis (SPEO) yang sudah ada dengan sistem yang menggunakan metode *Shallow NLP* dengan didukung oleh sinonim kata (Ade-Ibijola, dkk, 2012). Kemudian pada tahun 2014 metode Latent Semantic Analysis (LSA) dan Support Vector Machine (SVM) telah diimplementasikan untuk menghitung kemiripan teks dalam penilaian jawaban esai (Zhang, dkk, 2014). LSA merupakan metode untuk menentukan makna kata dan dokumen dengan menganalisis korpus teks yang besar. Dalam penelitiannya, mereka mendapatkan akurasi sebesar 88.8%. Masih pada tahun 2014, metode lain yang digunakan dalam penilaian jawaban esai adalah Regularized Latent Semantic Indexing (RLSI) terhadap bahasa China dengan tingkat akurasi 89% (Hao, dkk, 2014). Penelitian terus berlanjut, pada tahun 2015 telah menerapkan metode N-Gram untuk sistem penilaian jawaban esai pada bahasa Inggris dan menggunakan cosine similarity untuk menghitung kesamaan teks (Oduntan, dkk, 2015). Pada tahun yang sama terdapat penelitian penilaian esai otomatis menggunakan metode LSA dan penilaian *euclidean distance* (Yustiana, 2015). Pada tahun 2016, terdapat penelitian yang dibuat untuk fitur penilaian esai dimana didalamnya terdapat koreksi ejaan (Dong, 2016). Setiap sistem AES menggunakan beberapa teknik dasar atau hibrida teknik untuk evaluasi kebenaran esai. Teknik ini disebut teknik Natural Language Processing (NLP). NLP adalah bidang ilmu yang berkaitan dengan teknik, model dan algoritma yang digunakan dalam memproses dan menafsirkan esai. Mengklasifikasikan teknik untuk penilaian otomatis tanggapan teks bebas terdapat dalam tiga jenis utama: *Statistical*, *Information Extraction* dan *Full Natural Language Processing* (Mitchell, dkk, 2002). Pendekatan statistik, bila hanya berdasarkan analisis kata kunci, biasanya dianggap metode yang buruk, karena sulit untuk mengatasi masalah seperti sinonim atau polisemi dalam jawaban siswa, yang tidak

memperhitungkan urutan kata-kata dan tidak dapat menangani variabilitas leksikal. Di sisi lain, parsing teks lengkap dan semantic. Analisis sulit dilakukan, dan sangat sulit untuk pindah bahasa.

Information Extraction (IE), yang merupakan teknik yang digunakan dalam pekerjaan ini, berada di tengah statistik dan pendekatan NLP penuh. Ini hanya membutuhkan dasar NLP tanpa melakukan analisis mendalam dan ini lebih kuat daripada Analisis Kata Kunci biasa. Teknik IE berkaitan dengan perolehan informasi terstruktur dari teks bebas, misal mengidentifikasi Entitas Nama dalam teks dan mengisi template. IE dapat digunakan untuk mengekstrak dependensi antar konsep. Pertama, teks dipecah menjadi konsep dan hubungan mereka, maka dependensi yang ditemukan dibandingkan dengan pakar manusia untuk memberi nilai siswa. Misalnya, Automark dan ATM didasarkan pada pendekatan ini. Contoh dari skema penandaan yang diajukan oleh (Mitchell, dkk, 2002)

Full NLP adalah aplikasi metode komputasi untuk mengolah bahasa alami. Burstein, dkk, (2001) mengutip alat seperti parser sintaksis untuk menemukan struktur linguistik dari teks dan parser retorik untuk menemukan struktur wacana teks. Sebagai tambahan, Williams dan Dreher (2004) menggunakan *thesaurus* elektronik untuk mengekstrak informasi leksikal dan algoritma chunking yang dirancang khusus untuk mengekstrak frase kata benda dan klausa kata kerja. C-rater dan PS-ME (*Paperless School Marking Engine*) juga didukung oleh teknik-teknik ini. Kombinasi mereka meningkatkan penggunaan statistik dengan melibatkan parsing teks yang mendalam dan analisis semantik untuk mengumpulkan lebih banyak informasi untuk menilai secara efektif jawaban siswa.

K-means Clustering adalah algoritma pengelompokan iteratif sederhana dimana satu titik data hanya diperbolehkan berada dalam satu. Algoritma ini didasarkan pada minimisasi fungsi objektif yang menghitung ulang posisi centroid sampai lokasi mereka tidak lagi bergerak (Razon, dkk, 2010). Clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum (Tan, 2006).

Berdasarkan studi literatur yang dipaparkan sebelumnya, maka penulis akan mengimplementasikan sebuah sistem penilaian esai otomatis menggunakan *natural language processing* dikarenakan banyaknya metode NLP yang mempengaruhi tingkat akurasi dari sistem penilaian esai otomatis. NLP yang digunakan terdapat pembobotan *term frequency - inverse document frequency (tf-idf)* dan *n-gram*, yang didukung *WordNet* (kamus data). LSA memiliki tingkat akurasi yang cukup tinggi, sehingga diambil pembobotan TF-IDF dipilih karena salah satu dari metode LSA. N-Gram dipilih karena untuk mencari makna kalimat, sedangkan WordNet untuk mencari makna yang sama namun berbeda kata pada kalimat. Untuk tahap penilaiannya banyak tipe-tipe penilaian berdasarkan jarak kemiripan. Disini penulis menggunakan *clustering* dengan jenis *cluster* yaitu *K-Means*. *K-means* digunakan untuk mendapatkan nilai akhir yang diberikan sistem terhadap jawaban siswa. Dengan menerapkan *natural language processing* dan *K-means* diharapkan masalah penilaian ujian berbasis tes esai di atas dapat diselesaikan dengan pendekatan kajian komputasi ini. Selain itu, diharapkan juga adanya peningkatan akurasi dari penelitian sebelumnya. Hal ini dapat dilakukan dengan merancang dan membangun sistem penilaian yang mampu mengoreksi jawaban esai peserta tes secara otomatis.

1.2 Rumusan Masalah

Sesuai dengan apa yang telah dipaparkan pada latar belakang di sub bab sebelumnya, maka munculah rumusan masalah sebagai berikut:

1. Bagaimana merancang model sistem penilaian esai otomatis menggunakan *Natural Language Processing* dan *K-means*?
2. Bagaimana implementasi *Natural Language Processing* dan *K-means* terhadap sistem penilaian esai otomatis?
3. Bagaimana tingkat akurasi dari sistem penilaian esai otomatis yang dibangun dibandingkan dengan hasil penilaian manusia?

1.3 Batasan Masalah

Batasan Masalah dalam penelitian ini adalah sebagai berikut:

1. Sistem penilaian esai otomatis dikembangkan hanya untuk jawaban esai dalam bahasa Inggris yang tidak bersifat matematis, gambar, suara, ataupun video.
2. Soal esai yang digunakan merupakan soal yang memiliki jawaban esai yang menggunakan pertanyaan tanggapan terbatas.
3. Dalam penelitian ini jawaban siswa maupun kunci jawaban tidak menggunakan *Pronouns* atau kata ganti orang ketiga seperti “he”, “she”, “it”, “they”, “his”, “him”, dan lain – lain, jika terdapat data pronouns maka penulis akan ganti berdasarkan kata rujukannya.
4. Istilah-Istilah unsur kimia yang terdapat dalam jawaban siswa ataupun kunci jawaban guru yang berupa singkatan seperti “H”, “O”, “CO₂”, dan lain-lain, akan diakronimkan ke dalam bahasa Inggris menjadi “Hydrogen”, “Oxygen”, “Carbon Dioxide”, dan lain-lain.
5. Sistem penilaian esai otomatis ini ditujukan untuk sekolah-sekolah yang melakukan proses pembelajarannya dengan bilingual, terutama bahasa Inggris di Indonesia.

1.4 Tujuan Penelitian

Tujuan mengerjakan penelitian ini adalah sebagai berikut:

1. Merancang model penilaian soal esai otomatis dengan menggunakan pendekatan *Natural Language Processing* dan penilaian menggunakan K-means.
2. Mengimplementasikan model dari sistem penilaian esai otomatis pada tujuan penelitian pertama.
3. Mengetahui tingkat akurasi sistem penilaian jawaban esai otomatis yang dibangun dengan hasil penilaian manusia.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mempermudah pengajar dalam kegiatan penilaian evaluasi jawaban esai siswa.
2. Mempercepat waktu penilaian jawaban esai siswa.
3. Mengurangi inkonsistensi penilaian jawaban esai siswa oleh guru.

4. Guru dapat mengevaluasi kegiatan pembelajarannya berdasarkan hasil penilaian yang didapat.
5. Hasil dari penilaian dapat menjadi tolak ukur kemampuan siswa dalam memahami pelajaran.

1.6 Definisi Operasional

Sesuai dengan judul penelitian, terdapat istilah yang perlu diberikan penjelasan untuk menyamakan persepsi dalam penelitian ini. Penjelasan dalam penelitian ini adalah sebagai berikut:

1. Sistem Penilaian Esai

Sistem Penilaian Esai adalah sebuah teknologi komputer yang mengevaluasi dan menilai dari tulisan esai. (Shermis dan Burstein, 2003).

2. *Natural Language Processing*

Natural Language Processing (NLP) atau dalam bahasa Indonesia Pemrosesan Bahasa Alami, adalah cabang ilmu komputer dan linguistik yang mengkaji interaksi antara komputer dengan bahasa (alami) manusia (Charniak, 1984)

3. *WordNet*

WordNet (Kamus Data) adalah sebuah kumpulan atau wadah kata yang berhubungan dengan kesamaan makna atau sinonim dalam berbahasa Inggris.

4. *Term Frequency-Inverse Document Frequency (TF-IDF)*

Term frequency-inverse document frequency adalah statistik numerik yang dimaksudkan untuk mencerminkan betapa pentingnya sebuah kata dalam sebuah dokumen dalam kumpulan atau korpus.

5. *N-Gram*

N-Gram adalah potongan N karakter dalam suatu string atau potongan N kata dalam kalimat tertentu. Sehingga dengan menggunakan *N-Gram*, susunan kata dalam sebuah kalimat dapat diperhatikan.

6. *K-Means*

K-Means merupakan salah satu algoritma *clustering* (X Wu dan V Kumar, 2009). Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa

label kelas. Hal ini berbeda dengan supervised learning yang menerima masukan berupa vektor $(-x-1, y1)$, $(-x-2, y2)$, ..., $(-x-i, yi)$, dimana x_i merupakan data dari suatu data pelatihan dan y_i merupakan label kelas untuk x_i (Russel,2010).

1.7 Sistematika Penulisan

Pada bagian sistematika penulisan ini akan diuraikan mengenai penjelasan pada setiap bab.

BAB I PENDAHULUAN

Bab ini menjelaskan bagaimana penelitian dapat muncul dan mengenai konteks penelitian yang dilakukan, diawali dengan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, definisi operasional dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Pada bab ini dijelaskan teori pendukung untuk melakukan penelitian. Teori yang dijelaskan dalam bab ini yaitu mengenai, evaluasi pembelajaran, *automated essay scoring*, *natural language processing* termasuk didalamnya *text preprocessor*, *term frequency-inverse document frequency*, *n-gram*, *clustering*, dan *K-Means*.

BAB III METODE PENELITIAN

Bab ini menjelaskan langkah-langkah penelitian yang akan dilakukan, dimulai dari desain penelitian, fokus penelitian, alat dan bahan yang digunakan untuk penelitian dan metode penelitian.

BAB IV TEMUAN DAN PEMBAHASAN

Bab ini menjabarkan hasil penelitian yang telah dilakukan dan analisisnya. Semua pertanyaan mengenai masalah yang diangkat dalam tema skripsi dibahas di sini. Yaitu tentang proses pengumpulan data,

pengembangan model, implementasi sistem, studi kasus, desain eksperimen, dan analisa.

BAB V KESIMPULAN, IMPLIKASI DAN REKOMENDASI

Bab ini berisi kesimpulan, serta implikasi dan rekomendasi bagi peneliti selanjutnya dari hasil penelitian yang telah dilakukan.