

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penilaian adalah proses penghimpunan dan cara mengolah informasi untuk menentukan ukuran capaian dari hasil belajar peserta didik. Ada 3 unsur yang biasanya dievaluasi adalah pengetahuan (*knowledge*), keterampilan (*Skill*), dan sikap (*attitude*). Diantara beberapa bentuk penilaian tersebut pilihan ganda (*multiple choice*), *matching test*, tes uraian singkat, dan esai. Seorang guru, dosen atau penilai biasanya memberikan soal yang kemudian diujikan kepada peserta didik dalam bentuk soal pilihan ganda ataupun soal esai. Soal esai banyak digunakan sekolah agar dapat mengetahui pemahaman kritis siswa [1].

Penilaian esai dikatakan dapat mengevaluasi hasil dari latihan pembelajaran yang menantang dan prosedur penilaian ujian esai dapat memengaruhi daya ingat, pemahaman, dan integrasi konsep siswa. Sebagai hasilnya, ketika siswa mempelajari topik-topik pelajaran, guru dapat mengukur tingkat pemahaman mereka. Namun, hal ini masih menimbulkan masalah baru dalam meninjau jawaban esai dan uraian, yang memakan waktu jika dibandingkan dengan jawaban pertanyaan pilihan ganda. Selain itu, sistem penilaian jawaban objektif setiap guru mungkin berbeda, sehingga masalah ini perlu diselesaikan.

Penelitian Handayani, dkk [2] menggunakan LSA dengan memperhatikan kata sinonim dari kunci jawaban siswa untuk membantu melakukan penilaian jawaban soal esai secara otomatis. Penelitian tersebut menggunakan data tes esai soal mata pelajaran Pendidikan Pancasila dan Kewarganegaraan (PPKn) dengan 5 soal esai, 34 siswa dengan masing-masing memiliki 5 jawaban dan panjang jawaban esai terdiri dari satu hingga tiga kalimat. Didapatkan akurasi tertinggi adalah 84,35% termasuk dalam kategori baik. Kesimpulannya, penelitian ini mempertimbangkan sinonim kata dan jawaban siswa terbukti dapat meningkatkan akurasi penilaian dibandingkan tanpa mempertimbangkan sinonim [2].

Metode LSA tersebut masih kurang bagus untuk menentukan nilai akurasi, maka

dari itu dikembangkan menjadi metode *Generalized Latent Semantic Analysis (GLSA)*. GLSA yang menghasilkan n-gram dokumen dibandingkan matriks kata per dokumen dari LSA[3].

Lemantara dkk, [4] membuat prototipe dengan metode tampilan dan kesamaan teks yang lebih baik menggunakan Metode GLSA. Hasil perhitungan menunjukkan bahwa prototipe yang dikembangkan dalam penelitian ini memiliki nilai tampilan yang lebih baik dan tingkat kepuasan pengguna yang meningkat sebesar 6,12%. Dibuktikan dengan hasil kepuasan pengguna dari 74,44% menjadi 80,56%. Dengan demikian, prototipe sebelumnya dapat diperbaiki untuk mengidentifikasi kesamaan teks dalam penilaian esai otomatis dan mendeteksi plagiarisme[4].

Dalam penelitian Cahyani, dkk mengkombinasikan TF-IDF (*Term Frequency-Inverse Document Frequency*) dan Word2Vec. TF-IDF memberikan bobot pada setiap kata berdasarkan seberapa sering kata tersebut muncul dalam dokumen dan seberapa unik kata tersebut dalam koleksi dokumen. TF-IDF menghasilkan fitur yang khas, tetapi fitur tunggal yang dihasilkan memiliki arti yang kurang jelas. Hal ini, disebabkan oleh ketidakmampuan TF-IDF untuk memperhitungkan keterhubungan semantik dengan kata-kata lain dalam dokumen yang sama atau dalam seluruh korpus [5]. Dapat disimpulkan bahwasanya kombinasi ekstraksi fitur TF-IDF dan Word2Vec pada klasifikasi buku perpustakaan dengan SVM (*Support Vector Machine*) menghasilkan akurasi tertinggi sebesar 73%. Oleh karena itu, pada penelitian ini digabung ekstraksi fitur TF-IDF dan Word2Vec agar dapat memperkuat representasi fitur dari konten item hal ini karena Word2Vec merupakan ekstraksi fitur kata yang dapat memperhitungkan semantik antar kata dari korpus[5].

Word2Vec memiliki dua jenis model, yaitu *Skip-Gram* dan *Continuous Bag of Words (CBOW)*. Perbedaan antara *Skip-Gram* dan CBOW dapat ditemui pada struktur *Neural Network*. *Skip-Gram* mengambil sebuah kata sebagai masukan dan meminta *Neural Network* untuk memprediksi kata-kata yang umumnya muncul di sekitar kata masukan tersebut. Sedangkan, CBOW menggunakan *Neural Network* untuk menebak sebuah kata berdasarkan kata-kata yang mengelilingi kata yang sedang ditebak [6].

Pada penelitian ini, hasil modifikasi linguistik dari kunci jawaban dan jawaban siswa

menjadi sebuah vektor yang dapat digunakan untuk dimasukkan sebagai input pertama dengan memanfaatkan *deep learning*. Dalam representasi kata, Word2vec menggunakan *Neural Network* untuk menunjukkan kesamaan kontekstual dan semantik setiap kata masukan yang terdiri dari one hot encoded vectors. Hasil dari proses konseptual akan merepresentasikan hubungan antara satu kata dengan kata lainnya [7].

Dengan adanya hal itu, diperlukan Penilaian Jawaban Esai Otomatis dengan cara kerja yang digunakan dalam melakukan penilaian otomatis pada soal uraian ini menggunakan sistem mengecek kemiripan antara jawaban siswa dengan kunci jawaban guru, dengan menerapkan metode *Generalized Latent Semantic Analysis* (GLSA). Evaluasi akan diselesaikan menggunakan pendekatan *cosine similarity*, *jaccard similarity*, dan *dice similarity* serta menggabungkan *TF-IDF* dan *Word2Vec*. Keluaran dari sistem ini adanya kemiripan nilai antara kunci jawaban guru dengan hasil jawaban siswa. Lembar jawaban dari setiap siswa dan guru kemudian akan diikuti oleh penerapan teknik *Singular Value Decomposition* (SVD), suatu metode aljabar linier ditransformasikan ke dalam bentuk vektor, yang diwakili oleh matriks U, S, dan V. Hal ini menyatakan bahwa semakin besar nilai yang diperoleh, semakin dekat nilai kedua vektor yang dihasilkan.

Dalam penelitian ini penulis fokus untuk melakukan penilaian jawaban soal esai pendek atau uraian secara otomatis MAPEL IPA. Mayoritas proyek penelitian Sains (IPA) saat ini dilakukan secara manual, menggunakan kertas dan pensil, yang berarti proses penelitian membutuhkan waktu yang relatif lama. Hal ini dapat merugikan pengumpulan atau analisis data jangka panjang, karena kualitas data yang dikumpulkan mungkin tidak valid dan konsisten. Oleh karena itu, diperlukan teknik penelitian baru yang mudah untuk memahami proyek penelitian Sains saat ini. Penelitian ini memilih mata pelajaran IPA karena soal esai IPA menuntut pemahaman konsep yang mendalam dan penjelasan rinci, sehingga proses penilaiannya lebih kompleks dibandingkan mata pelajaran lain. Selain itu, otomatisasi penilaian di IPA dapat lebih mudah diukur secara objektif, berbeda dengan mata pelajaran lain yang cenderung lebih subjektif atau memiliki jawaban pasti.[8].

Data yang digunakan dalam penelitian ini diambil pada nilai *Aessment* Harian IPA

kelas VII tahun ajaran 2024/2025 di MTS Negeri 1 Tarate Sumenep dengan 60 jumlah responden dan 20 soal. Hal ini dimaksudkan bahwa dengan menerapkan sistem penilaian esai otomatis ini dapat membantu guru dalam memberikan rekomendasi penilaian dan mengurangi jumlah waktu yang diperlukan untuk mengevaluasi jawaban siswa.

1.2 Rumusan Masalah

Berdasarkan pada uraian latar belakang maka dapat dipaparkan rumusan masalah sebagai berikut:

1.2.1 Permasalahan

Penilaian esai seringkali menjadi tugas yang memakan waktu dan rentan terhadap objektivitas. Penilaian jawaban esai otomatis yang memerhatikan variasi urutan kata dan pendekatan nyatanya belum bisa diatasi menggunakan LSA sehingga mempengaruhi perbandingan kesamaan dokumen jawaban siswa dan guru. Disisi lain, diperlukan *human score* untuk menilai kualitas esai. Namun, dengan kemajuan teknologi, metode penilaian otomatis semakin diminati. Salah satu metode yang digunakan adalah GLSA untuk menilai esai dengan cara yang lebih objektif dan konsisten.

1.2.2 Pertanyaan Penelitian

1. Bagaimana nilai RMSE (*Root Mean Square Error*) dari metode GLSA dengan perhitungan jarak *Cosine Similarity*, *Dice Similarity*, *Jaccard Similarity* dan parameter *N-gram* (*Unigram*, *Bigram*, *Trigram*) menggunakan ekstraksi fitur TF-IDF untuk menentukan penilaian esai otomatis *Assessment IPA*?
2. Bagaimana nilai RMSE dalam dari metode GLSA dengan perhitungan jarak *Cosine Similarity*, *Dice Similarity*, *Jaccard Similarity* dan parameter *N-gram* (*Unigram*, *Bigram*, *Trigram*) menggunakan ekstraksi fitur *word2vec* pada penilaian esai otomatis *Assessment IPA*?

1.3 Tujuan

1. Mengetahui nilai RMSE pada metode GLSA dengan perhitungan jarak *Cosine Similarity*, *Dice Similarity* atau *Jaccard Similarity* dan parameter *N-gram* (*Unigram*, *Bigram*, *Trigram*) menggunakan ekstraksi fitur TF-IDF untuk menentukan penilaian esai otomatis *Assessment IPA*

2. Mengetahui nilai RMSE pada metode GLSA dengan perhitungan jarak *Cosine Similarity*, *Dice Similarity* atau *Jaccard Similarity* dan parameter N-gram (Unigram, Bigram, Trigram) menggunakan ekstraksi fitur *word2vec* pada penilaian esai otomatis *Assessment IPA*

1.4 Manfaat

Manfaat yang diharapkan dari penelitian ini meliputi peningkatan performa penilaian esai otomatis, mengembangkan alat bantu pembelajaran, mengurangi beban kerja guru dengan menyediakan sistem penilaian esai otomatis yang lebih objektif serta membantu meningkatkan kualitas dan efisiensi pembelajaran dalam penilaian di tingkat pendidikan.

1.5 Batasan Masalah

1. Dataset yang digunakan mencakup bobot penilaian, nilai guru , serta informasi mengenai jawaban esai siswa pada mata pelajaran Ilmu Pengetahuan Alam (IPA) di MTs Negeri 1 Tarate Sumenep. Terdapat total 20 pertanyaan dan jawaban dari 60 siswa dalam dataset ini.
2. Mengevaluasi data tekstual, tidak mengevaluasi respons yang disajikan dalam bentuk suara, gambar, atau rumus matematika.
3. Model Word2Vec yang digunakan ialah model *Skip-gram*.
4. Memanfaatkan *Stopwords Removal* dari pustaka Sastrawi dan pustaka *Stemming* dari Nazief Adriani dan menggunakan *library nltk* dan *sastrawi*
5. Pengujian yang dilakukan menggunakan pengujian RMSE.
6. Fokus pada otomatisasi penilaian esai pendek, bukan esai panjang

1.6 Sistematika kepenulisan

Sistematika penulisan skripsi ini adalah sebagai berikut:

Bab I berisi tentang pendahuluan yang dilakukan pada penelitian ini. Bab ini membahas terkait latar belakang permasalahan yang ada, tujuan dan manfaat dilakukannya untuk penelitian, batasan yang ada dalam penelitian, serta isi tentang sistematika penulisan proposal.

Bab II berisi tentang kajian pustaka yang menjelaskan mengenai materi dan metode yang dijadikan landasan teori dalam penelitian ini. Di antaranya materi yang

dibahas adalah algoritma *Generalized Latent Semantic Analysis* dan penelitian terkait.

Bab III membahas mengenai Metode Usulan berisi penjelasan tentang *dataset* yang akan digunakan, alur kerja sistem, tahapan pengujian, skenario uji coba.

Bab IV membahas tentang implementasi serta pengujian berisi implementasi dari metode (*Generalized Latent Semantic Analysis*) GLSA untuk memecahkan permasalahan yang kemudian dilaksanakan pengujian untuk mengukur hasilnya.

Bab V berisi kesimpulan dari implementasi yang sudah dilaksanakan, dan juga saran agar dapat mengembangkan penelitian berikutnya.