

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini bertujuan untuk melakukan analisis komprehensif terhadap dua pendekatan dalam sistem *Automatic Short Answer Scoring* (ASAS), yaitu *direct scoring* dan *similarity-based scoring*, pada dua skenario evaluasi yang berbeda, yaitu *specific-prompt* dan *cross-prompt*. Analisis dilakukan secara kuantitatif melalui metrik evaluasi, seperti RMSE dan *Pearson Correlation*, analisis residual untuk mengidentifikasi outlier, analisis performa pada level soal, dan analisis kualitatif pada soal tertentu.

Hasil penelitian menunjukkan bahwa pendekatan *direct scoring* secara konsisten unggul dibandingkan *similarity-based scoring*, baik pada skenario *specific-prompt* dan *cross-prompt*. Keunggulan ini ditandai dengan nilai RMSE yang rendah, yang menunjukkan kesalahan prediksi yang lebih kecil, korelasi *Pearson* yang tinggi, yang menandakan hubungan linear antara skor prediksi dan skor asli itu kuat, dan jumlah outlier yang lebih sedikit pada ambang batas 0.1 maupun 0.2, yang mencerminkan kestabilan model dalam memprediksi skor. Hasil ini menunjukkan bahwa penggunaan representasi vektor secara langsung memberikan informasi semantik yang lebih kaya, dibandingkan hanya mengukur kesamaan semantic antar kalimat seperti pendekatan *similarity-based scoring*.

Namun, hasil analisis performa pada level soal menunjukkan bahwa *similarity-based scoring* memiliki keunggulan pada soal yang memiliki jawaban referensi tetap atau eksplisit, seperti soal yang meminta siswa untuk menuliskan isi teks proklamasi atau menjabarkan singkatan. Di sisi lain, *direct scoring* lebih fleksibel dalam menangani berbagai karakteristik soal, baik soal deskriptif yang memungkinkan variasi luas dalam jawaban siswa, maupun soal dengan jawaban referensi tetap yang memiliki struktur lebih beragam. Selain itu, melalui analisis kualitatif didapat bahwa kedua pendekatan memiliki keterbatasan dalam menilai jawaban yang secara konteks terlihat benar tetapi isi informasinya salah. Hal ini menunjukkan bahwa kedua pendekatan belum robust menghadapi jawaban yang mengandung misinformasi.

5.2 Saran

Berdasarkan hasil penelitian, terdapat dua saran yang dapat diberikan untuk pengembangan selanjutnya. Pertama, perlu menambahkan variasi jawaban pada dataset, khususnya jawaban yang secara struktur atau konteks tampak benar tetapi mengandung informasi yang salah. Hal ini penting karena hasil analisis kualitatif menunjukkan kedua pendekatan masih memberikan skor yang tinggi pada jawaban seperti ini. Kedua, disarankan untuk mengeksplorasi penggunaan model verifikasi fakta sebelum proses penilaian. Model ini berfungsi untuk memeriksa kebenaran isi jawaban dengan membandingkannya terhadap referensi atau pengetahuan eksternal, sehingga sistem dapat menghasilkan penilaian yang lebih akurat dan dapat diandalkan. Selain itu, disarankan untuk mengeksplorasi kedua pendekatan ini pada dataset berbahasa lain, serta menganalisis hubungan antara karakteristik soal dengan efektivitas pendekatan. Jika terbukti bahwa pendekatan tertentu lebih optimal untuk karakteristik soal tertentu, maka pengembangan sistem adaptif yang dapat mengenali karakteristik soal dan memilih pendekatan yang sesuai menjadi arah penelitian selanjutnya.