

BAB V

SIMPULAN & SARAN

5.1 Simpulan

Berdasarkan penelitian yang telah dilakukan, maka didapatkan kesimpulan seperti berikut:

1. Implementasi model berbasis *Vision Transformer* dilakukan dengan mengekstraksi fitur melalui bagian *encoder*, yang menghasilkan representasi dalam bentuk *embedding*. Proses *pooling* yang digunakan adalah mean *pooling*, yaitu pengambilan rata-rata dari seluruh vektor token. Pendekatan ini efektif dalam mereduksi *noise* pada data sekaligus menangkap informasi yang lebih menyeluruh dari keseluruhan input. Hasilnya, *embedding* yang dihasilkan menjadi lebih berkualitas dan mampu merepresentasikan fitur secara lebih baik. Hal ini terlihat dari hasil visualisasi pada tahap *initial clustering*, di mana gambar-gambar yang tergolong dalam satu kluster menunjukkan kemiripan, seperti contohnya sama-sama mengandung tabel, memiliki jumlah baris yang serupa, atau *layout* yang mirip.
2. Hasil eksperimen menunjukkan bahwa meskipun ketiga model mendapatkan hasil yang buruk pada tahap *initial clustering*, namun model berbasis *Vision Transformer* (ViT) mengalami kenaikan performa yang signifikan bila dibandingkan dengan model berbasis *Convolutional Neural Network* (CNN). Hasil tersebut memberikan kesimpulan bahwa model berbasis ViT lebih unggul dibandingkan model berbasis CNN. Model ViT lebih baik dalam mempelajari data baru yang diluar dari domain data latihnya dibandingkan model berbasis CNN (Lowe dkk., 2024). Hal ini menyebabkan ViT memberikan fitur yang lebih representatif dibandingkan CNN ketika selesai melalui proses *DeepCluster*, yang selanjutnya fitur tersebut membantu memberikan hasil *clustering* yang lebih baik. Terlihat pula bahwa setelah melalui proses iterasi lebih banyak yakni 10 kali iterasi, model CNN masih belum mampu untuk memberikan hasil *clustering* sebaik model ViT.

3. Berdasarkan hasil eksperimen terlihat bila suatu kelompok terbentuk ketika memiliki persamaan pola visual seperti contoh bentuk tabel yang terkandung dalam jawaban yang menjadi ciri khas suatu kelompok hal ini membuat ketika terdapat gambar tabel dalam jawabannya maka gambar jawaban esai itu akan ada pada kelompok yang sama. Selanjutnya beberapa ciri dalam satu kelompok adalah seperti bentuk resolusi gambar yang sama, jarak antara baris tulisan yang rapat atau renggang, dan latar kertas menjadi suatu ciri khas yang menentukan gambar jawaban tersebut berada dalam satu kelompok.
4. Hasil metrik evaluasi pada tahap *initial clustering* untuk setiap model menunjukkan hasil yang kurang baik, terlihat dari kecilnya angka dari masing-masing metrik evaluasi. Hal ini menandakan klaster yang terbentuk masih belum membentuk kelompok yang padat dan juga terpusat. Hal ini disebabkan karena representasi fitur yang dihasilkan masih belum cukup baik sehingga data masih tersebar pada ruang dimensi tinggi. Dapat dilihat dari visualisasi, distribusi klaster cenderung masih tersebar secara acak dan juga sangat renggang yang memberikan indikasi bahwa fitur belum cukup baik untuk menghasilkan klaster yang berkualitas baik. Selanjutnya setelah melalui proses *DeepCluster* dimana dalam prosesnya terjadi *fine-tuning* terhadap model menggunakan *pseudo labels* yang dihasilkan pada tahap *initial clustering*, model menjadi lebih baik dalam merepresentasikan fitur, hal ini terlihat dari hasil visualisasi yang mulai menunjukkan distribusi yang mulai terpusat dan berada pada wilayahnya masing-masing. Nilai pada setiap metrik evaluasi *final clustering* menunjukkan kenaikan yang cukup signifikan yang menandakan bahwa kualitas *clustering* meningkat seiring meningkatnya hasil representasi fitur yang dihasilkan model.
5. Hasil visualisasi menggunakan algoritma seperti t-SNE yang terkadang memberikan hasil yang kurang sejalan dengan nilai evaluasi metrik. Hasil visualisasi yang terlihat terpusat dan padat seringkali memiliki nilai evaluasi metrik yang rendah. Hal ini dapat terjadi karena pada sejatinya perhitungan nilai evaluasi metrik seperti SIL, DBI, dan CHI dilakukan pada ruang dimensi asli atau ruang dimensi tinggi, berbeda dengan t-SNE yang berada pada ruang dua dimensi. Visualisasi jarak yang dihasilkan t-SNE merupakan perhitungan

probabilistik yang menjadikan jarak yang tampak pada plot bukan jarak asli pada ruang dimensi tinggi. Dengan demikian, hasil visualisasi dari algoritma t-SNE menjadi gambaran kualitatif mengenai seberapa baik representasi fitur data, dan apakah data cenderung dapat membentuk klaster yang terpisah.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, berikut beberapa saran yang dapat dijadikan sebagai pertimbangan untuk penelitian selanjutnya:

1. Model seperti CNN menunjukkan hasil visualisasi yang lebih rapih akan tetapi menunjukkan hasil nilai metrik yang lebih rendah dibandingkan model ViT, hal ini dapat terjadi karena besarnya dimensi dari fitur yang dihasilkan oleh model CNN. Oleh karena itu, dapat dilakukan teknik pengurangan dimensi yang efisien agar fitur yang akan dijadikan input algoritma *clustering* dapat lebih mudah diproses, sehingga hasil nilai metrik evaluasi dapat meningkat.
2. Pada penelitian ini konteks kemiripan hanya berdasarkan pola visual dari gambar jawaban esai mahasiswa tanpa mempertimbangkan konteks isi jawaban yang ada. Penelitian selanjutnya dapat mengeksplorasi pendekatan yang mempertimbangkan isi konteks jawaban esai mahasiswa, seperti melakukan transkripsi terhadap jawaban lalu dilakukan *clustering* terhadap jawaban tersebut