

6. KESIMPULAN DAN SARAN

6.1. Kesimpulan

Berikut merupakan kesimpulan dari pengerjaan skripsi setelah melalui proses implementasi dan pengujian :

1. Pada rumusan masalah nomor 1, nilai metrik ROUGE dan BERTScore dari hasil ringkasan model PGN + BERTSum-*Extractive*-IndoBERT yaitu ROUGE-1 adalah 53,15%, ROUGE-2 adalah 43,32%, ROUGE-L adalah 51,10%, dan BERTScore adalah 81,05%.
2. Pada rumusan masalah nomor 2, karena hasil ringkasan PGN bersifat ekstraktif, tidak ada *hallucination* pada ringkasan PGN. Meskipun IndoBERT bersifat ekstraktif, namun terdapat beberapa kasus dimana IndoBERT dapat menyebabkan *hallucination* dengan menghasilkan kata atau frasa yang tidak ada pada *input* (ringkasan PGN). *Error* ini dapat terjadi ketika IndoBERT berusaha melakukan *text generation*. Maka dari itu, PGN + BERTSum-*Extractive*-IndoBERT belum dapat memperbaiki *out-of-context error*.
3. Pengaplikasian *pretrained model* IndoBERT dapat mengoreksi *syntactical error* pada ringkasan PGN. Kata atau frasa repetitif yang gagal ditangani *coverage mechanism* dari PGN, dapat dikoreksi pada ringkasan final oleh IndoBERT. IndoBERT juga dapat mengoreksi frasa yang salah dengan melakukan *text generation* kata atau frasa yang dibutuhkan untuk mengoreksi frasa. Pada kondisi tertentu, IndoBERT dapat *generate* kata atau frasa penghubung, untuk menggabungkan 2 kalimat menjadi 1 kalimat, atau sebagai penghubung antar kalimat. Pada sisi lain, *pretrained model* IndoBERT tidak dapat mengeliminasi kalimat yang tidak relevan dengan kalimat lain.
4. Perubahan panjang *minimum* dan *maximum* pada hasil ringkasan PGN dan IndoBERT tidak memberi pengaruh besar pada metrik skor ROUGE maupun BERTScore. Namun, karena hasil ringkasan cenderung mengambil beberapa kalimat pertama pada artikel, ringkasan yang sedikit lebih panjang dapat meliputi lebih banyak informasi, terutama pada beberapa kondisi dimana terdapat informasi penting di tengah artikel.
5. Eksperimen dengan mengubah *word embedding* dengan menggunakan *pretrained model* IndoBERT belum dapat memperoleh skor yang lebih baik dibandingkan dengan Tensorflow *word embedding*.
6. Dari hasil survei dengan 3 *sample* artikel dan hasil *generated summary* yang tidak terdapat *error*, sebagian besar responden menganggap hasil ringkasan sudah cukup baik dari segi interpretabilitas, relevansi, serta panjang ringkasan.

6.2. Saran

Berikut merupakan beberapa saran untuk melanjutkan atau mengembangkan penelitian serupa :

1. Untuk model *Pointer-Generator Networks*, dapat menambahkan dataset dengan ringkasan referensi yang lebih abstrak. Hal ini dapat meningkatkan performa *generator* pada PGN.
2. Melakukan *hyperparameter tuning* pada model *Pointer-Generator Networks*, dan melakukan *fine tuning* pada model IndoBERT.
3. Menerapkan *transformers* sebagai *Encoder* pada model *Pointer-Generator Networks*.