

Tentu, sebagai rekan sejawat (sesama dosen dan pegiat AI/Data Science), saya akan memberikan review yang objektif, tajam, namun konstruktif terhadap draft paper berjudul "**Revitalizing the Endangered Papua Kokas Language: A Low-Resource NMT Approach using Transformers**" ini.

Secara umum, topiknya sangat menarik dan mulia (preservasi bahasa daerah). Namun, secara teknis (terutama dari sudut pandang *Machine Learning* dan metodologi penelitian), ada beberapa **kelemahan mendasar (fundamental weaknesses)** yang cukup fatal jika tidak diperbaiki sebelum *submit* ke jurnal bereputasi (seperti JISEBI yang dituju).

Berikut adalah poin-poin kritiknya:

1. Pemilihan Pre-trained Model yang Kurang Tepat (Fundamental)

Ini adalah isu teknis terbesar di paper ini.

- **Temuan:** Pada bagian *Model Configuration*, disebutkan: "*In this study, we fine-tuned the Helsinki-NLP/opus-mt-id-en pre-trained model...*"¹.
- **Kritik:** Anda menggunakan model yang dilatih untuk menerjemahkan Bahasa Indonesia ke **Bahasa Inggris** (id-en).
 - Decoder dari model ini memiliki *vocabulary* dan *embedding space* yang dioptimalkan untuk Bahasa Inggris.
 - Memaksa decoder Bahasa Inggris untuk memproduksi Bahasa Papua Kokas (yang secara morfologi dan kosa kata pasti sangat berbeda) hanya dengan **2.908 kalimat** adalah langkah yang sangat tidak efisien. Model harus melakukan "catastrophic forgetting" terhadap bahasa Inggris untuk belajar bahasa baru dari nol.
- **Saran:** Seharusnya gunakan model **Multilingual** yang memang didesain untuk *transfer learning* ke bahasa baru (Many-to-Many atau Many-to-One), seperti **mBART, NLLB (No Language Left Behind)**, atau setidaknya model MarianMT yang multilingual (opus-mt-mul-en atau sejenisnya yang punya *vocabulary* lebih luas). Penggunaan tokenizer bahasa Inggris untuk bahasa Papua juga berisiko menghasilkan tokenisasi yang buruk (memecah kata menjadi subword yang tidak bermakna).

2. Skor BLEU yang "Too Good To Be True" (Indikasi *Data Leakage*)

- **Temuan:** Paper melaporkan skor BLEU mencapai **45-46²²²²²²²²²²**.
- **Kritik:** Dalam dunia NMT, apalagi *low-resource* (hanya ~2.900 data), skor BLEU di atas 40 itu sangat jarang terjadi kecuali ada masalah pada data. Biasanya skor 20-30 sudah dianggap bagus untuk *low-resource*.
 - **Dugaan Data Leakage:** Mengingat data diambil dari "daily conversational expressions"³, besar kemungkinan ada kalimat yang sangat mirip atau duplikat antara *Training Set* dan *Validation Set*.

- Contoh: Jika di training ada "Saya makan nasi" dan di test ada "Saya makan roti", model mungkin hanya menghafal pola, bukan "belajar" menerjemahkan (generalisasi).
- **Saran:** Cek ulang metode *splitting* data. Pastikan tidak ada *overlap* kalimat yang signifikan. Lakukan analisis *Lexical Overlap* antara train dan test set. Jika datanya sangat repetitif, skor BLEU tinggi itu semu (bias).

3. Absennya Baseline Pembanding

- **Temuan:** Paper mengklaim Transformer lebih baik dari RNN/LSTM berdasarkan literatur⁴, tetapi **tidak melakukan eksperimen perbandingan sendiri**.
- **Kritik:** Dalam paper eksperimental, Anda tidak bisa hanya mengklaim metode A bagus tanpa membandingkannya dengan metode B pada dataset yang sama.
- **Saran:** Anda wajib melatih model sederhana (misalnya LSTM/GRU standard atau Statistical Machine Translation/SMT seperti Moses) pada dataset 2.908 kalimat tersebut. Jika Transformer (yang berat) ternyata performanya sama saja atau hanya beda tipis dengan SMT (yang ringan) pada data sekecil ini, maka penggunaan Transformer menjadi kurang terjustifikasi secara komputasi.

4. Inkonsistensi Data & Validasi

- **Inkonsistensi Split Ratio:**
 - Di teks⁵ tertulis rasio **80:20**.
 - Di Tabel 1⁶ tertulis **Training (90%)** dan **Validation (10%)**.
 - Ini kesalahan fatal yang menunjukkan ketidaktelitian penulis.
- **Validasi yang Lemah:** Hanya mengandalkan BLEU⁷. Untuk bahasa daerah yang strukturnya unik, BLEU seringkali tidak merepresentasikan akurasi makna. Tanpa evaluasi manusia (walaupun paper mengakui ini sebagai limitasi⁸), validitas hasil terjemahan sangat diragukan. Minimal berikan analisis error (error analysis) secara manual pada output model, bukan hanya menampilkan contoh yang "bagus-bagus" saja di Tabel 4.

5. Masalah Referensi (Editorial)

- **Temuan:** Bagian REFERENSI⁹ masih berisi *placeholder* atau *template* bawaan IEEE (misalnya: "Title of paper if known," *unpublished*¹⁰, "We suggest there should be at least 20 references"¹¹).
- **Kritik:** Ini menunjukkan draft ini belum siap sama sekali secara editorial. Mengirimkan

paper dengan daftar pustaka *template* akan langsung ditolak (desk reject) oleh editor jurnal.

Kesimpulan & Rekomendasi Langkah Selanjutnya

Paper ini memiliki potensi, tetapi **metodologi eksperimennya perlu diperbaiki**. Sebagai sesama dosen di bidang ini, saran saya:

1. **Ganti Base Model:** Coba fine-tune model multilingual (misal: facebook/nllb-200-distilled-600M atau mBART).
2. **Buat Baseline:** Latih model RNN sederhana atau SMT sebagai pembanding.
3. **Audit Data:** Pastikan pembagian train/test benar-benar bersih dari duplikasi.
4. **Perbaiki Referensi:** Hapus template dan masukkan referensi riil yang relevan.

Apakah Anda ingin saya bantu buatkan **skema arsitektur/diagram alur** perbaikan untuk metodologi penelitiannya agar lebih kuat saat revisi?