

**PENDEKATAN GENERATIF BERBASIS MODEL BAHASA
PRALATIH UNTUK EKSTRAKSI TRIPLET OPINI**

Laporan Tugas Akhir

Disusun sebagai syarat kelulusan tingkat sarjana

**Oleh
William
NIM : 13518138**



**PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
Juni 2022**

**PENDEKATAN GENERATIF BERBASIS MODEL BAHASA
PRALATIH UNTUK EKSTRAKSI TRIPLET OPINI**

Laporan Tugas Akhir

Oleh

William

NIM : 13518138

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung

Telah disetujui dan disahkan sebagai Laporan Tugas Akhir
di Bandung, pada tanggal 22 Juni 2022

Pembimbing,

Dr. Masayu Leylia Khodra, ST., MT

NIP19760429 200812 2 001

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Pengerjaan dan penulisan Laporan Tugas Akhir ini dilakukan tanpa menggunakan bantuan yang tidak dibenarkan.
2. Segala bentuk kutipan dan acuan terhadap tulisan orang lain yang digunakan di dalam penyusunan laporan tugas akhir ini telah dituliskan dengan baik dan benar.
3. Laporan Tugas Akhir ini belum pernah diajukan pada program pendidikan di perguruan tinggi mana pun.

Jika terbukti melanggar hal-hal di atas, saya bersedia dikenakan sanksi sesuai dengan Peraturan Akademik dan Kemahasiswaan Institut Teknologi Bandung bagian Penegakan Norma Akademik dan Kemahasiswaan khususnya Pasal 2.1 dan Pasal 2.2.

Bandung, 22 Juni 2022



William

NIM 13518138

ABSTRAK

PENDEKATAN GENERATIF BERBASIS MODEL BAHASA

PRALATIH UNTUK EKSTRAKSI TRIPLET OPINI

Oleh

WILLIAM

NIM : 13518138

Ekstraksi triplet opini bertujuan untuk mendapatkan daftar triplet ekspresi aspek, ekspresi sentimen, dan polaritas sentimen yang terkandung dalam ulasan. Ekstraksi triplet opini eksplisit telah dipelajari secara ekstensif. Namun sayangnya masih belum ada penelitian yang meneliti ekstraksi triplet opini dengan aspek implisit. Salah satu metode untuk melakukan ekstraksi triplet opini adalah dengan pendekatan pembangkitan teks yang bertujuan untuk membangkitkan triplet opini secara langsung. Keunggulan pendekatan tersebut adalah keluaran model tidak terbatas pada masukan model yang dapat membantu penanganan triplet opini dengan aspek implisit. Penelitian Tugas Akhir ini berfokus untuk melakukan ekstraksi triplet opini dengan pendekatan pembangkitan teks, serta penanganan aspek implisit dalam ekstraksi triplet opini.

Ekstraksi triplet opini dengan pendekatan pembangkitan teks dapat dilakukan dengan modifikasi konstruksi korpus data latih pada dokumen ulasan sesuai dengan format anotasi *extraction-style*, sehingga daftar triplet opini teks ulasan dapat langsung dibangkitkan sesuai dengan format *extraction-style*. Penanganan terhadap triplet opini dengan aspek implisit dapat dilakukan dengan menganotasikan aspek spesial NULL pada triplet opini tersebut. Tugas Akhir menggunakan model bahasa pralatih IndoT5 untuk membangkitkan triplet opini dari teks ulasan.

Berdasarkan hasil eksperimen, konfigurasi terbaik model ekstraksi triplet opini pada Tugas Akhir adalah dengan model bahasa pralatih IndoT5, penggunaan jarak levenshtein dalam *postprocess* dan k (jumlah kandidat) adalah 1. Berdasarkan pengujian untuk ekstraksi triplet opini eksplisit, model tugas akhir berhasil melampaui kinerja model *baseline*, yaitu model adaptasi GTS pada Tugas Akhir Wirawan. Model Tugas Akhir mendapat nilai *F1-score* 0.8117 untuk ekstraksi triplet opini eksplisit. Sementara itu, pengujian untuk ekstraksi triplet opini dengan penanganan aspek implisit mendapatkan nilai *F1-score* 0.7950.

Kata kunci: ASBA, ekstraksi triplet opini, pendekatan pembangkitan teks.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa berkat petunjuk serta rahmat-Nya penulis dapat menyelesaikan karya Tugas Akhir ini yang berjudul “Pendekatan Generatif Berbasis Model Bahasa Pralatih untuk Ekstraksi Triplet Opini”. Penulis juga ingin mengucapkan terima kasih kepada pihak-pihak lain yang telah membantu penulis dalam memberi saran dan dukungan selama pengerjaan Tugas Akhir ini terutama kepada:

1. Ibu Dr.Masayu Leylia Khodra, ST, MT. selaku dosen pembimbing yang sabar dan telah membimbing, memberi arahan, saran, masukan, dan dukungan lainnya kepada penulis dalam mengerjakan tugas akhir ini.
2. Bapak Ir. Windy Gambetta, M.B.A. dan Ibu Dr. Fariska Zakhralativa Ruskanda, S.T., M.T. selaku dosen penguji yang memberikan evaluasi dan masukan dalam pengerjaan tugas akhir.
3. Staf pengajar dan tata usaha Program Studi Teknik Informatika yang senantiasa membantu mengurus proses administrasi tugas akhir.
4. Keluarga penulis yang selalu memberikan dukungan moral. Khususnya kedua orang tua penulis yang selalu percaya pada penulis bahkan saat penulis tidak percaya pada dirinya sendiri.
5. Seluruh pihak lain yang telah membantu penulis baik langsung maupun tidak langsung yang tidak dapat disebutkan satu per satu.

Sebagai akhir kata, penulis menyadari bahwa tugas akhir ini memiliki banyak kekurangan. Untuk itu, penulis akan selalu menerima dengan senang hati segala kritik dan saran yang dapat memperbaiki tugas akhir ini. Penulis berharap agar tugas akhir yang telah penulis susun ini dapat bermanfaat bagi pembaca dan dapat berkontribusi terhadap pemajuan ilmu terutama di bidang Informatika.

DAFTAR ISI

ABSTRAK	iv
DAFTAR ISI.....	vi
DAFTAR LAMPIRAN	ix
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN.....	1
I.1 Latar Belakang.....	1
I.2 Rumusan Masalah.....	3
I.3 Tujuan	3
I.4 Batasan Masalah	4
I.5 Metodologi.....	4
I.6 Sistematika Pembahasan.....	5
BAB II STUDI LITERATUR	7
II.1 Ekstraksi Triplet Opini.....	7
II.2 Ekstraksi Triplet Opini dengan Pendekatan Pembangkitan Teks	8
II.3 Metode Pengukuran Jarak pada Teks	12
II.3.1 Jarak Levenshtein.....	12
II.3.2 <i>Cosine Similarity</i> menggunakan Vektor Representasi Kata	13
II.4 Tokenisasi	13
II.5 Model Bahasa Pralatih Berbasis Transformer	15
II.6 Evaluasi Kinerja Ekstraksi Triplet Opini.....	18

II.7	Penelitian Terkait.....	19
II.7.1	Penelitian Wirawan (2021)	19
II.7.2	Penelitian Genadi (2021).....	21
II.7.3	Penelitian Rendyanto (2021).....	23
BAB III Analisis Masalah dan Perancangan Solusi		26
III.1	Analisis Persoalan	26
III.2	Analisis Solusi.....	28
III.2.1	Konstruksi Korpus dengan Format <i>Extraction-style</i>	28
III.2.2	Model Bahasa Pralatih	29
III.2.3	Langkah <i>Postprocess</i>	30
III.2.4	Anotasi Triplet Opini dengan Aspek Implisit.....	31
III.3	Rancangan Solusi	32
III.3.1	Modul Praproses	34
III.3.2	Modul <i>Fine-Tune</i> Model Ekstraksi Triplet Opini.....	34
III.3.3	Modul <i>Posttraining</i>	36
III.3.4	Modul Pembangkitan Triplet Opini.....	37
III.3.5	Modul <i>Postprocess</i>	38
III.4	Desain Eksperimen.....	40
III.5	Implementasi	41
BAB IV Eksperimen dan pengujian.....		46
IV.1	Eksperimen	46
IV.1.1	Tujuan Eksperimen	46
IV.1.2	Data.....	46

IV.1.3	Skenario Eksperimen	48
IV.2	Hasil Eksperimen	51
IV.2.1	Hasil Eksperimen dengan Dataset Eksplisit	51
IV.2.2	Hasil Eksperimen dengan Dataset Campuran.....	55
IV.3	Analisis Hasil Pengujian dengan Dataset Eksplisit	57
IV.4	Analisis Hasil Pengujian dengan Dataset Campuran	62
IV.5	Analisis Pengaruh <i>Posttraining</i> pada Ekstraksi Triplet Opini	64
BAB V	KESIMPULAN DAN SARAN	66
V.1	Kesimpulan	66
V.2	Saran	67
DAFTAR REFERENSI		68

DAFTAR LAMPIRAN

Lampiran A. Perbaikan Anotasi Triplet Opini.....	70
A.1 Perbaikan Anotasi Polaritas Triplet Opini	70
A.2 Perbaikan Anotasi Pasangan Triplet Opini	72
A.3 Penambahan Anotasi Pasangan Triplet Opini dengan Aspek Implisit	73
Lampiran B. Detail Cara Kerja Model Berbasis Transformer	74
B.1 Pelatihan Model	74
B.2 Inferensi Model.....	75

DAFTAR GAMBAR

Gambar II.1. Rincian Hasil Eksperimen Zhang dkk., (2021)	11
Gambar II.2 Pseudocode Algoritma Jarak Levenshtein (Jurafsky dan Martin; 2021)	12
Gambar II.3 Contoh Perhitungan Jarak Levenshtein (Jurafsky dan Martin; 2021)	13
Gambar II.4 Arsitektur Transformer (Vaswani, dkk., 2017)	16
Gambar II.5 Ilustrasi Objektif Pelatihan Tanpa Supervisi dari Proses Pre-training T5 (Raffel, dkk., 2020).....	16
Gambar II.6 Ilustrasi Jenis Triplet (Zhang, dkk., 2020).....	19
Gambar II.7 Arsitektur Sistem Ekstraksi Triplet Opini pada Tugas Akhir Wirawan (2021)	20
Gambar II.8 Arsitektur Sistem Ekstraksi Triplet Opini Berbasis Span pada Tesis Genadi (2021).....	22
Gambar II.9 Arsitektur Sistem Klasifikasi Kategori Aspek dan Klasifikasi Polaritas Sentimen Tugas Akhir Rendyanto (2021).....	24
Gambar III.1. Arsitektur Sistem Ekstraksi Triplet Opini dengan Pendekatan Pembangkitan Teks	33
Gambar III.2 Visualisasi Kerja Modul Fine-Tune Model Ekstraksi Triplet Opini	35
Gambar III.3 Ilustrasi Proses <i>Post-training</i> untuk IndoT5	36
Gambar III.4 Ilustrasi Proses Pembangkitan Triplet Opini.....	37
Gambar III.5 Ilustrasi Proses Perbaikan Prediksi dengan <i>Postprocess</i>	39
Gambar III.6 Detail Implementasi Solusi	41

DAFTAR TABEL

Tabel II.1 Konfigurasi Model Bahasa Pralatih IndoT5.....	17
Tabel II.2 Hasil Kinerja Terbaik Model Tugas Akhir Wirawan (2021)	21
Tabel II.3 Contoh Kegagalan Ekstraksi Triplet Opini pada Sentimen yang Terletak di Depan Aspek (Wirawan, 2021).....	21
Tabel II.4 Hasil Kinerja Terbaik Model Tugas Akhir Rendyanto (2021).....	25
Tabel III.1. Konstruksi Korpus dengan Format <i>Extraction-style</i>	28
Tabel III.2 Contoh Hasil Inferensi Model Bahasa Pralatih.....	29
Tabel III.3 Contoh Anotasi Data.....	31
Tabel III.4 Contoh Anotasi Triplet Opini dengan Aspek Implisit	32
Tabel III.5 Contoh Masukan dan Keluaran Modul Pengolahan Teks.....	34
Tabel III.6 Contoh Masukan dan Keluaran Modul Pembangkitan Triplet Opini .	38
Tabel IV.1 Detail Dataset Triplet Opini Eksplisit.....	47
Tabel IV.2 Detail Dataset Triplet Opini Campuran	47
Tabel IV.3 Skenario Eksperimen	49
Tabel IV.4 Konfigurasi Varian Model untuk Dataset Eksplisit dan Campuran ...	50
Tabel IV.5 Nilai Hiperparameter Eksperimen	51
Tabel IV.6 Konfigurasi <i>Post-training</i> IndoT5	51
Tabel IV.7 Kinerja Validasi Skenario P0.....	51
Tabel IV.8 Kinerja Validasi Skenario P1.....	52
Tabel IV.9 Contoh Kasus <i>Typo</i> Hasil <i>Postprocess</i>	53
Tabel IV.10 Contoh Kasus Bukan <i>Typo</i> Hasil <i>Postprocess</i>	53
Tabel IV.11 Kinerja Validasi Skenario P2.....	54

Tabel IV.12 Detail Kinerja Variasi Model pada Dataset Eksplisit	55
Tabel IV.13 Kinerja Validasi Skenario P0.....	56
Tabel IV.14 Kinerja Validasi Skenario P1	56
Tabel IV.15 Kinerja Variasi Skenario P2	56
Tabel IV.16 Detail Kinerja Variasi Model pada Dataset Campuran	57
Tabel IV.17 Hasil Pengujian Model Terbaik Dibandingkan Model Baseline untuk Ekstraksi Triplet Opini Eksplisit.....	58
Tabel IV.18 Ringkasan Hasil Pengujian Model.....	59
Tabel IV.19 Contoh Kasus Kesalahan Umum 1	60
Tabel IV.20 Contoh Kasus Kesalahan Umum 2	60
Tabel IV.21 Contoh Kasus Kesalahan Akibat Sedikit Perbedaan pada Ekspresi dalam Triplet Opini	61
Tabel IV.22 Contoh Kasus Kesalahan pada Kalimat Saran atau Pernyataan	61
Tabel IV.23 Hasil Pengujian Model Terbaik untuk Ekstraksi Triplet Opini dengan Aspek Implisit	62
Tabel IV.24 Contoh Kasus Kesalahan Umum 1 Triplet Opini Aspek Implisit	63
Tabel IV.25 Contoh Kasus Kesalahan Umum 2 Triplet Opini Aspek Implisit	63
Tabel IV.26 Perbedaan Representasi Sebelum dan Sesudah <i>Post-training</i>	64
Tabel IV.27 Contoh Kasus Kesalahan Akibat <i>Posttraining</i>	65

BAB I

PENDAHULUAN

I.1 Latar Belakang

Teks ulasan merupakan suatu bentuk evaluasi yang dapat dimanfaatkan untuk mengetahui opini konsumen terhadap suatu hal yang berguna dalam mengambil keputusan yang tepat sasaran serta relevan dengan pendapat konsumen. Sebagai contoh, “kamarnya **sangat bersih**, tapi pegawainya **kurang sopan**” menunjukkan bahwa konsumen merasa fasilitas hotel telah baik dan perlu dipertahankan, namun perlu peningkatan pada kesopanan pegawai. Bidang studi yang menganalisis sentimen suatu opini terhadap suatu entitas beserta atributnya dinamakan analisis sentimen atau penggalian opini (Liu, 2015).

Ekstraksi triplet opini merupakan *task* turunan dari analisis sentimen berbasis aspek yang bertujuan untuk mendapatkan semua pasangan triplet opini (ekspresi aspek, ekspresi sentimen, dan polaritas sentimen) dari teks ulasan (Peng, dkk., 2019; Wu, dkk., 2020). Ekstraksi triplet opini dilakukan untuk mengetahui maksud ulasan yang diberikan secara utuh mulai dari apa yang sedang dibicarakan dalam ulasan (ekspresi aspek), bagaimana tanggapan pemberi ulasan terhadap target opini tersebut (polaritas sentimen) dan mengapa pemberi ulasan memberikan respon demikian (ekspresi sentimen) terhadap target opini terkait (Peng, dkk., 2019).

Ekstraksi triplet opini untuk domain perhotelan bahasa Indonesia telah dilakukan oleh Wirawan (2021) dan Genadi (2021). Tugas Akhir Wirawan (2021) memiliki kinerja terbaik untuk ekstraksi triplet opini pada ulasan domain perhotelan. Meskipun demikian, ulasan produk juga mengandung banyak opini dengan aspek yang tidak eksplisit. Kebanyakan dari studi yang dilakukan mengabaikan masalah ini (Cai, dkk., 2021).

Penelitian Zhang dkk., (2022) juga mengutarakan pada sebagian besar kumpulan data *benchmark* analisis sentimen berbasis aspek yang tersedia, terdapat aspek (entitas) dirujuk secara implisit dalam teks ulasan karena diasumsikan telah diketahui oleh pemberi ulasan. Sebagai contoh, pada ulasan “baik dan bersih, namun pelayan cerewet”, pemberi ulasan mengasumsikan aspek “kamar hotel” telah diketahui sehingga tidak ditulis. Ekstraksi triplet opini eksplisit hanya akan mendapatkan informasi mengenai pelayan yang cerewet. Padahal, opini terkait hotel yang bagus dan bersih yang dirujuk secara implisit dalam ulasan juga mengandung informasi.

Penelitian Wirawan (2021) melakukan ekstraksi triplet opini dengan mengadaptasi pendekatan *sequence labelling* dengan skema pelabelan berbasis *grid* (GTS) (Wu, dkk., 2020) untuk melabeli setiap relasi pasangan token yang kemudian dipasangkan secara berbasis aturan untuk mendapatkan setiap pasangan triplet opini. Salah satu batasan yang ada dalam penelitian tersebut adalah hanya mengekstraksi pasangan triplet opini eksplisit. Hal ini karena pendekatan *sequence labelling* pada skema pelabelan berbasis *grid* (GTS) sulit diterapkan pada triplet opini dengan aspek implisit karena tidak ada token yang dapat dilabeli untuk mendapatkan aspek implisit tersebut.

Ekstraksi triplet opini dapat dilakukan dengan pendekatan pembangkitan teks. Zhang dkk., (2021) mengusulkan pendekatan pembangkitan triplet opini secara langsung untuk menangani kasus aspek implisit dengan bantuan model bahasa pralatih. Eksperimen pada penelitian tersebut menunjukkan bahwa kinerja model dengan pendekatan pembangkitan teks lebih baik dari model lainnya pada 4 dataset: data ulasan laptop (SemEval 2014) dan 3 data ulasan restoran (SemEval 2014, SemEval 2015, dan SemEval 2016) yang menunjukkan hasil penelitian Zhang dkk., (2021) merupakan *state-of-the-art* dibanding penelitian lainnya untuk ekstraksi triplet opini.

Model bahasa dapat menetapkan probabilitas bersyarat untuk setiap kemungkinan token berikutnya dengan menghitung distribusi dari seluruh kosakata yang ada pada

model bahasa. Hal ini dapat mengatasi permasalahan yang ada dalam pendekatan *sequence labelling* dalam menangani triplet opini dengan aspek implisit karena keluaran model tidak terbatas hanya pada masukan model. Model bahasa yang digunakan pada penelitian Zhang dkk., (2021) adalah model bahasa pralatih T5 untuk membangkitkan pasangan triplet opini dari teks masukan. T5 sendiri merupakan salah satu model bahasa pralatih berbasis transformers yang diusulkan Raffel dkk., (2020) yang bertujuan untuk mengubah permasalahan pemrosesan bahasa menjadi teks ke teks.

I.2 Rumusan Masalah

Walaupun sudah terdapat penelitian terkait ekstraksi triplet opini eksplisit dan sudah terdapat dataset pasangan triplet opini eksplisit untuk teks bahasa Indonesia. Namun, penelitian terkait ekstraksi triplet opini dengan penanganan terhadap aspek implisit berbasiskan model bahasa pralatih untuk teks bahasa Indonesia belum ada. Rumusan masalah pada tugas akhir ini adalah sebagai berikut.

1. Bagaimana cara membangun model ekstraksi triplet opini aspek eksplisit dan implisit untuk teks bahasa Indonesia dengan pendekatan pembangkitan teks dengan memodifikasi penelitian Zhang dkk., (2021)?
2. Bagaimana perbandingan kinerja model yang diusulkan dalam penelitian ini dengan model penelitian Wirawan (2021) dalam ekstraksi triplet opini eksplisit?

I.3 Tujuan

Tujuan utama dari penelitian tugas akhir ini adalah sebagai berikut.

1. Menghasilkan model ekstraksi triplet opini aspek eksplisit dan implisit dengan pendekatan pembangkitan teks.
2. Membandingkan kinerja model yang dihasilkan dengan model *baseline* dari penelitian sebelumnya pada ekstraksi triplet opini eksplisit.

I.4 Batasan Masalah

Batasan masalah pada penelitian tugas akhir ini adalah dataset yang digunakan berisi teks ulasan domain perhotelan berbahasa Indonesia sebagai bahasa utama.

I.5 Metodologi

Metodologi pengerjaan penelitian tugas akhir ini dapat dijabarkan sebagai berikut.

1. Analisis Permasalahan dan Solusi

Pada tahap ini dilakukan analisis persoalan yang dihadapi untuk menyelesaikan permasalahan ekstraksi triplet opini dengan pendekatan pembangkitan teks. Kemudian dilakukan studi literatur terhadap alternatif solusi-solusi yang dapat digunakan untuk menangani persoalan-persoalan yang ada.

2. Konstruksi Korpus

Pada tahap ini dilakukan konstruksi korpus yang digunakan untuk membangun model ekstraksi triplet opini dengan pendekatan pembangkitan teks. Korpus yang digunakan adalah korpus pada Tugas Akhir Wirawan (2021) dengan format anotasi yang disesuaikan untuk pelatihan pendekatan pembangkitan teks.

3. Desain Eksperimen

Pada tahap desain eksperimen dilakukan perancangan arsitektur solusi serta penentuan tujuan, skenario dan skema eksperimen yang akan dilakukan. Selain itu, tahap ini juga dilakukan perencanaan pembangunan modul-modul yang dibutuhkan untuk eksperimen.

4. Implementasi

Pada tahap ini dilakukan pembangunan modul-modul yang diperlukan dalam eksperimen, yaitu modul untuk pelatihan model beserta modul utilitas untuk menjalankan skenario eksperimen yang telah didefinisikan.

5. Eksperimen

Pada tahap ini dilakukan eksperimen untuk menguji efektifitas penggunaan pendekatan pembangkitan teks dengan menggunakan model bahasa pralatih dalam ekstraksi triplet opini serta untuk mencari konfigurasi terbaik. Selain itu, dilakukan analisis terhadap hasil eksperimen tersebut.

I.6 Sistematika Pembahasan

Penulisan laporan tugas akhir ini akan dikelompokkan menjadi beberapa bab dengan tujuan agar mudah dipahami dengan jelas sebagai berikut.

BAB I PENDAHULUAN

Bab I berisi tentang latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi, dan sistematika pembahasan penelitian tugas akhir. Latar belakang bab ini membahas mengenai batasan penelitian ekstraksi triplet opini saat ini, pemanfaatan pendekatan pembangkitan teks untuk bisa mengatasi batasan tersebut serta gambaran umum hal yang dikerjakan pada tugas akhir ini.

BAB II STUDI LITERATUR

Bab II berisikan teori yang berkaitan dengan topik yang dibahas pada tugas akhir. Teori yang dibahas mencakup penjelasan mengenai ekstraksi triplet opini, penjelasan tentang ekstraksi triplet opini dengan pendekatan pembangkitan teks, model bahasa pralatih dan teknik evaluasi yang digunakan pada penelitian tugas akhir. Selain itu, bab II juga berisi beberapa studi dan referensi mengenai penelitian terkait untuk ekstraksi triplet opini dan analisis sentimen berbasis aspek dengan pendekatan pembangkitan teks.

BAB III IDENTIFIKASI MASALAH DAN RANCANGAN SOLUSI

Bab III menjelaskan mengenai analisis persoalan yang dihadapi secara lebih rinci, analisis solusi yang akan dilakukan dalam tugas akhir dan rancangan solusi dalam membangun model ekstraksi triplet opini dengan penanganan aspek implisit. Selain itu, dibahas juga desain eksperimen yang dilakukan untuk mendapatkan konfigurasi terbaik model dan penjelasan mengenai implementasi masing-masing modul dalam tugas akhir.

BAB IV EKSPERIMEN DAN EVALUASI HASIL

Bab IV menjelaskan metode yang tujuan eksperimen, detail data serta detail skenario eksperimen yang akan diterapkan yang akan diterapkan pada tugas akhir ini. Selain itu, bab IV juga menjabarkan perbandingan kinerja hasil eksperimen ekstraksi triplet opini eksplisit dengan hasil penelitian Wirawan (2021) beserta analisis terhadap hasil eksperimen dan evaluasi dari solusi tugas akhir untuk ekstraksi triplet opini dengan aspek implisit.

BAB V PENUTUP

Bab V berisi simpulan dan saran yang berkaitan dengan analisis dan evaluasi terhadap model yang telah dibangun serta saran untuk pengembangan selanjutnya.

BAB II

STUDI LITERATUR

II.1 Ekstraksi Triplet Opini

Masalah analisis sentimen umumnya terdiri dari dua komponen utama: target dan sentimen (Zhang, dkk., 2022). Target dapat berupa entitas atau aspek yang dimiliki entitas, sedangkan sentimen dapat berupa opini positif/netral/negatif yang diungkapkan tentang target. Pada analisis sentimen berbasis aspek, target dapat dideskripsikan dengan kategori aspek (c) atau ekspresi aspek (a), sedangkan sentimen melibatkan ekspresi opini terperinci yaitu ekspresi opini (o) dan polaritas sentimen (p) (Zhang, dkk., 2022).

Pada analisis sentimen berbasis aspek, entitas juga dapat dianggap sebagai sebuah aspek umum khusus dalam sebuah opini, sehingga secara kolektif juga dapat disebut sebagai aspek (Zhang, dkk., 2022). Suatu aspek dapat disebutkan secara eksplisit dalam suatu tinjauan atau tersirat melalui ungkapan lain (implisit). Misalnya, layar dalam kalimat “Layar lebar” adalah aspek eksplisit. Sebaliknya, ekspresi aspek implisit sering diidentifikasi melalui kata sifat (Su, dkk., 2008). Jelas, jika aspek muncul secara eksplisit di unit ulasan, maka aspek bersifat eksplisit. Sebaliknya, aspek bersifat implisit apabila aspek tidak ditulis secara eksplisit dan aspek dapat disimpulkan dengan kata-kata opini di sekitarnya dalam ulasan (Su, dkk., 2008).

Ekstraksi triplet opini bertujuan untuk melakukan ekstraksi ekspresi aspek, ekstraksi ekspresi sentimen dan klasifikasi polaritas sentimen untuk aspek tersebut serta ekstraksi relasi untuk pemasangan kedua ekspresi dalam menghasilkan triplet opini (a, o, p). *Task* ini pertama kali diajukan pada penelitian Peng dkk., (2019) untuk menyelesaikan seluruh permasalahan utama dalam analisis sentimen berbasis aspek dengan membentuk satu *task* gabungan yang dapat memberikan informasi lengkap terkait suatu ulasan mulai dari **apa** aspek yang ditinjau, **bagaimana**

polaritas sentimen terhadap aspek tersebut dan **mengapa** aspek tersebut memiliki polaritas tersebut (Peng, dkk., 2019).

Contoh dari ekstraksi triplet opini dapat dilihat sebagai berikut. Triplet opini dari teks ulasan “Pelayan sangat ramah tetapi makanan kurang lezat” adalah (“Pelayan”, “ramah”, “POS”) dan (“makanan”, “kurang lezat”, “NEG”) (Peng, dkk., 2019). Dari kedua triplet opini tersebut dapat diketahui bahwa aspek pelayan memiliki polaritas sentimen positif dengan alasan pelayan tersebut ramah, sedangkan untuk aspek makanan memiliki polaritas sentimen negatif dengan alasan makanan kurang lezat.

Dalam ekstraksi triplet opini eksplisit untuk bahasa Indonesia, penelitian yang telah dilakukan menggunakan pendekatan klasifikasi teks dan *sequence labelling*. Penelitian sebelumnya seperti penelitian pada tugas akhir Wirawan (2021) yang mengadaptasi pendekatan skema pelabelan *grid* (GTS) telah berhasil membangun model ekstraksi triplet opini dengan pendekatan *sequence labelling*. Pendekatan klasifikasi teks juga dapat dilakukan seperti pada penelitian tesis Genadi (2021) yang mengadaptasi *framework* SpanMLT untuk melakukan klasifikasi teks pada representasi *span* guna mendapatkan triplet opini dari kalimat ulasan.

II.2 Ekstraksi Triplet Opini dengan Pendekatan Pembangkitan Teks

Penelitian terkait analisis sentimen berbasis aspek umumnya memerlukan proses pembangunan model secara spesifik untuk menyesuaikan masukan dan keluaran dari masing-masing tugas dalam analisis sentimen berbasis aspek. Zhang dkk., (2021) mengusulkan *Generative Aspect-based Sentiment Analysis* (GAS) yang bertujuan untuk membentuk suatu pendekatan umum yang dapat menyelesaikan berbagai permasalahan analisis sentimen berbasis aspek dengan pendekatan pembangkitan teks termasuk ekstraksi triplet opini.

Sebelumnya, pendekatan pembangkitan teks untuk analisis sentimen berbasis aspek telah dilakukan pada penelitian Rendyanto (2021) dengan penambahan *auxiliary sentence* pada akhir dokumen ulasan untuk menyelesaikan *task* kategorisasi aspek

dan klasifikasi polaritas sentimen. Hal ini memungkinkan model bahasa pralatih untuk digunakan secara langsung tanpa proses pelatihan (*zero shot*). Solusi ini dapat diterapkan untuk *task* kategorisasi aspek dan klasifikasi polaritas sentimen karena keluaran yang diharapkan cukup pendek dan terdiri dari kumpulan kosa kata yang telah didefinisikan dari awal. Solusi ini sulit diterapkan pada permasalahan yang lebih kompleks seperti ekstraksi triplet opini karena model perlu membangkitkan seluruh triplet opini dengan format tertentu untuk bisa mengekstraksi tiap komponen dalam triplet tersebut dengan tepat.

Agar ekstraksi triplet opini yang umumnya dilakukan dengan pendekatan *sequence labelling* atau klasifikasi teks dapat dilakukan dengan pendekatan pembangkitan teks dengan menggunakan dua format anotasi korpus untuk anotasi data latih model agar masukan dan keluaran model berbentuk teks. Kedua format anotasi korpus tersebut dinamakan *annotation-style* dan *extraction-style*.

Perbedaan kedua format tersebut dapat dilihat pada anotasi pasangan triplet opini dari kalimat ulasan “The Unibody construction is solid, sleek and beautiful”. Pada *extraction-style* target keluaran model adalah triplet opini yaitu “(Unibody construction, solid, positive); (Unibody construction, sleek, positive); (Unibody construction, beautiful, positive)”, sementara pada *annotation-style* target keluaran model adalah teks masukan beserta informasi tambahan berupa triplet opini yaitu “The [Unibody construction | positive | solid, sleek, beautiful] is solid, sleek and beautiful”.

Proses *decoding* kedua format anotasi juga sedikit berbeda untuk mengekstrak triplet opini hasil bangkitan model. Pada *extraction-style*, dilakukan ekstraksi hasil keluaran dengan memisahkan masing-masing triplet opini dengan pemisah yang sudah ditetapkan di awal (“;”). Sementara pada *annotation-style*, dilakukan proses ekstraksi hasil keluaran yang terdapat dalam kurung kurawal (“[]”). Hasil penelitian Zhang dkk., (2021) menunjukkan penggunaan *extraction-style* lebih cocok digunakan untuk *task* kompleks seperti ekstraksi triplet opini dibandingkan *annotation-style*.

Korpus kemudian digunakan untuk melatih model bahasa pralatih T5 (Raffel, dkk., 2020) sebagai model bahasa untuk membangkitkan triplet opini. Setelah pelatihan dilakukan, triplet opini yang diinginkan dapat diperoleh dengan melakukan mengekstrak hasil keluaran model (y') sesuai dengan format anotasi korpus yang digunakan pada data latih.

Proses ekstraksi dari hasil keluaran model dilakukan dengan melakukan proses *decoding* pada y' . Sebagai contoh untuk *extraction-style*, dilakukan ekstraksi konten yang termasuk dalam tanda kurung “()” dari y' dan pemisahan elemen triplet opini (a, o, p) yang berbeda dengan koma “,” sebagai pemisah. Apabila proses ekstraksi gagal dilakukan karena terjadi kesalahan format pada y' maka prediksi tersebut diabaikan.

Penanganan ketika teks yang dibangkitkan berbeda dari target (tidak terdapat dalam teks masukan) dapat dilakukan dengan suatu langkah normalisasi prediksi. Idealnya, hasil bangkitan triplet opini dari model seharusnya ada pada teks masukan. Namun, hal ini belum tentu benar karena model membangkitkan teks berdasarkan kumpulan kosa kata model pralatih. Untuk menangani hal tersebut diusulkan langkah normalisasi prediksi untuk membandingkan dan memperbaiki tiap elemen triplet opini yang dibangkitkan model dengan teks masukan.

Langkah normalisasi prediksi yang diusulkan dilakukan dengan membentuk kumpulan kosa kata untuk tiap teks masukan. Khusus untuk kumpulan kosa kata polaritas sentimen hanya mengandung kosa kata dari polaritas yang memungkinkan (positif, negatif dan netral). Sedangkan untuk ekspresi aspek dan ekspresi sentimen, kumpulan kosa kata yang dimaksud adalah semua token dari teks masukan. Kemudian dilakukan pengecekan terhadap keluaran triplet opini dengan kumpulan kosa kata terkait. Jika terdapat keluaran yang tidak sesuai, maka akan dicari kosa kata paling mirip dalam kumpulan kosa kata yang diukur dengan jarak levenshtein.

Varian arsitektur dari model ekstraksi triplet opini yang diajukan dievaluasikan terhadap beberapa dataset, yaitu data ulasan laptop (L14) dan restoran (R14, R15, dan R16) dari Semeval 2014 Task 4, Semeval 2015 Task 12 dan Semeval 2016

Task 5. Dataset-dataset tersebut disediakan oleh penelitian Fan dkk., (2019), Wang dkk., (2017), dan Wang dkk., (2016).

	L14	R14	R15	R16
CMLA+ (Wang et al., 2017)	33.16	42.79	37.01	41.72
Li-unified-R (Li et al., 2019a)	42.34	51.00	47.82	44.31
Pipeline (Peng et al., 2020)	42.87	51.46	52.32	54.21
Jet (Xu et al., 2020)	43.34	58.14	52.50	63.21
Jet+BERT (Xu et al., 2020)	51.04	62.40	57.53	63.83
GAS-ANNOTATION-R	52.80	67.35	56.95	67.43
GAS-EXTRACTION-R	<u>58.19</u>	<u>70.52</u>	60.23	<u>69.05</u>
GAS-ANNOTATION	54.31	69.30	<u>61.02</u>	68.65
GAS-EXTRACTION	60.78	72.16	62.10	70.10

Gambar II.1. Rincian Hasil Eksperimen Zhang dkk., (2021)

Rincian hasil dari eksperimen tersebut dapat dilihat pada Gambar II.1. Dalam eksperimen tersebut juga ditampilkan hasil prediksi tanpa *postprocess* yang ditunjukkan dengan akhiran “-R”. Hasil tersebut adalah nilai *F1-score* rata-rata dari 5 kali eksperimen dengan nilai *seed* yang berbeda. GAS dengan konfigurasi model terbaik berhasil mendapatkan *F1-score* yang lebih unggul pada semua dataset *benchmark* untuk ekstraksi triplet opini.

Dari hasil eksperimen yang dilakukan, disimpulkan bahwa langkah normalisasi prediksi yang diterapkan bermanfaat dalam menangani kasus ketika model membangkitkan kata dengan perbedaan leksikal yang minor dari target seperti pada kata “*repeat*” dengan target “*repeats*” serta kasus ketika model membangkitkan kata yang memiliki turunan yang sama dengan target seperti “*vegan*” dengan target “*vegetarian*”. Namun, Zhang dkk., (2021) mengamati strategi normalisasi umumnya gagal pada kasus ketika model membangkitkan kata yang cukup berbeda secara leksikal dari target seperti pada kata “*salmon*” dengan target “*spinach*”.

II.3 Metode Pengukuran Jarak pada Teks

II.3.1 Jarak Levenshtein

Jarak levenshtein (atau sering disebut *edit distance*) adalah suatu ukuran kesamaan antara dua teks. Jarak yang dimaksud adalah jumlah langkah yang diperlukan untuk mengubah teks asal menjadi teks target. Langkah yang dapat dilakukan adalah memasukkan, menghapus, atau mengganti karakter dalam teks asal. Semakin rendah jarak levenshtein antara 2 teks, maka semakin serupa kedua teks tersebut. Algoritma untuk mengukur jarak levenshtein dapat dilihat pada Gambar II.2.

Sebagai contoh, jarak antara kata "*intention*" dan "*execution*" adalah 5 operasi (hapus karakter "i", substitusi karakter "e" dengan "n", substitusi "x" dengan "t", memasukkan karakter "c", substitusi karakter "u" dengan "n"). Visualisasi jarak levenshtein antara 2 kata tersebut dapat dilihat pada Gambar II.3.

```
function MIN-EDIT-DISTANCE(source, target) returns min-distance
    n ← LENGTH(source)
    m ← LENGTH(target)
    Create a distance matrix  $D[n+1, m+1]$ 

    # Initialization: the zeroth row and column is the distance from the empty string
     $D[0,0] = 0$ 
    for each row  $i$  from 1 to  $n$  do
         $D[i,0] \leftarrow D[i-1,0] + \text{del-cost}(\text{source}[i])$ 
    for each column  $j$  from 1 to  $m$  do
         $D[0,j] \leftarrow D[0,j-1] + \text{ins-cost}(\text{target}[j])$ 

    # Recurrence relation:
    for each row  $i$  from 1 to  $n$  do
        for each column  $j$  from 1 to  $m$  do
             $D[i,j] \leftarrow \text{MIN}( D[i-1,j] + \text{del-cost}(\text{source}[i]),$ 
                                $D[i-1,j-1] + \text{sub-cost}(\text{source}[i], \text{target}[j]),$ 
                                $D[i,j-1] + \text{ins-cost}(\text{target}[j]) )$ 

    # Termination
    return  $D[n,m]$ 
```

Gambar II.2 Pseudocode Algoritma Jarak Levenshtein (Jurafsky dan Martin; 2021)

i n t e n t i o n	← delete i
n t e n t i o n	← substitute n by e
e t e n t i o n	← substitute t by x
e x e n t i o n	← insert u
e x e n u t i o n	← substitute n by c
e x e c u t i o n	

Gambar II.3 Contoh Perhitungan Jarak Levenshtein (Jurafsky dan Martin; 2021)

II.3.2 *Cosine Similarity* menggunakan Vektor Representasi Kata

Cosine similarity adalah ukuran kesamaan antara 2 teks yang mengukur jarak sudut kosinus antara dua vektor representasi teks yang diproyeksikan dalam bidang multidimensi (Sitikhu, dkk., 2019). Artinya, jarak didapatkan berdasarkan perbedaan terhadap orientasi vektor dan besar vektor (Sitikhu, dkk., 2019). *Cosine similarity* dapat diukur dengan persamaan II-1 (Sitikhu, dkk., 2019), dimana vektor A dan B adalah dua vektor representasi numerik dari teks yang ingin dicari jaraknya. *Cosine similarity* sering digunakan dalam bidang pemrosesan bahasa untuk mengukur seberapa mirip dua kata/kalimat berdasarkan vektor representasi numeriknya.

$$similarity = \cos(\theta) = \frac{A \cdot B}{|A||B|} \quad (II-1)$$

Word Embeddings adalah representasi vektor dari kata yang diperoleh dengan melatih model pada korpus besar dengan pelatihan tanpa supervisi (Sitikhu, dkk., 2019). Teknik ini telah banyak digunakan dalam klasifikasi teks menggunakan kesamaan semantik. Vektor yang didapat dari *word embedding* tersebut dapat digunakan untuk mencari jarak antara 2 teks yang ingin dicari jaraknya. (Sitikhu, dkk., 2019).

II.4 Tokenisasi

Tokenisasi merupakan proses pemisahan frasa, kalimat, paragraf atau dokumen menjadi unit yang lebih kecil seperti per kata. Masing-masing unit yang lebih kecil ini disebut sebagai token (Jurafsky dan Martin; 2021). Sebagai contoh, pada kalimat

“Saya akan pergi ke hotel tersebut” dapat dipisah menjadi [“Saya”, “akan”, “pergi”, “ke”, “hotel”, “tersebut”] dengan tokenisasi per kata. Proses tokenisasi ini bermanfaat untuk mempersiapkan teks sebagai masukan dari model sebagai bagian dari tahapan praproses teks.

Meskipun terdapat banyak algoritma yang dapat digunakan untuk menerapkan proses tokenisasi, namun ide dasar dari proses tokenisasi relatif serupa yaitu untuk memisahkan teks menjadi bagian yang lebih kecil. Dalam tokenisasi per kata, algoritma yang paling sering digunakan yaitu memecah kalimat berdasarkan spasi seperti yang contoh pada paragraf sebelumnya. Selain tokenisasi per kata, juga terdapat tokenisasi berbasis karakter yang akan memecah kalimat menjadi karakter terpisah. Sebagai contoh, teks “Saya akan pergi ke hotel tersebut” akan dipisah menjadi [“S”, “a”, “y”, “a”, “a”, “k”, “a”, “n”,..., “s”, “e”, “b”, “u”, “t”]. Teknik tokenisasi karakter ini efektif untuk bahasa yang tidak menggunakan spasi seperti bahasa Mandarin dan Jepang.

Selain kedua teknik tokenisasi teks tersebut, terdapat opsi lain dalam tokenisasi. Alih-alih mendefinisikan pemisah token dalam proses tokenisasi ataupun algoritma sejenis untuk memisahkan teks menjadi token, proses tokenisasi dapat langsung menggunakan data latih yang telah tersedia untuk menentukan secara otomatis token yang diperlukan dalam kumpulan kosa kata *tokenizer* berdasarkan subkata dalam data latih. Hal ini bermanfaat dalam mengatasi permasalahan terkait kata yang tidak dikenal (OOV), mengingat kasus OOV merupakan salah satu masalah utama dalam pemrosesan bahasa (Jurafsky dan Martin; 2021).

Untuk menangani permasalahan OOV, *tokenizer* modern umumnya menggunakan kumpulan kosa kata yang mencakup token yang lebih kecil dari sebuah kata, yang sering disebut sebagai subkata. Setiap kata yang belum pernah dilihat *tokenizer* dapat dipisah menjadi subkata yang telah dikenali *tokenizer*, seperti kata “menginap” dapat dipisah menjadi “meng” dan “-inap”. Pada umumnya, skema tokenisasi modern terdiri dari 2 tahapan yaitu pembelajaran token dan segmentasi token. Tahap pembelajaran token dilakukan dengan menggunakan korpus data latih

untuk membentuk kumpulan kosa kata *tokenizer* yang terdiri dari subkata-subkata korpus data latih. Sedangkan tahap segmentasi token dilakukan pada korpus data uji untuk dapat memproses teks data uji menjadi subkata yang dikenali *tokenizer*.

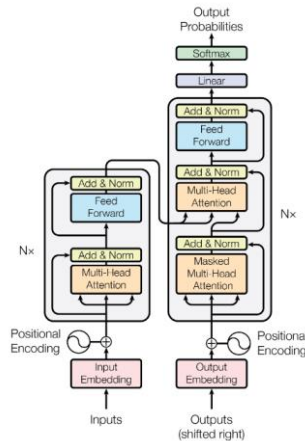
Terdapat 3 (tiga) macam algoritma tokenisasi subkata yang sering digunakan yaitu *byte-pair encoding* atau sering disebut BPE, *unigram language modelling*, dan *WordPiece*. Kemudian terdapat SentencePiece yang digunakan dalam model bahasa pralatih T5. SentencePiece merupakan *tokenizer* dan *de-tokenizer* subkata untuk pemrosesan bahasa alami. Sentencepiece mendukung implementasi dari *byte-pair encoding* dan *unigram language modelling* untuk melakukan segmentasi subkata dan dapat mengubah subkata tersebut menjadi sekuens id untuk digunakan model (Kudo dan Richardson, 2018).

II.5 Model Bahasa Pralatih Berbasis Transformer

Transformer adalah salah satu arsitektur model yang mengeliminasi koneksi rekuren dalam menangani data bertipe sekuens sehingga tidak memerlukan penggunaan RNN maupun konvolusi. Metode ini menggunakan mekanisme atensi sebagai mekanisme utama dalam pembentukan arsitekturnya. Arsitektur ini dapat menangani perkara sekuens dengan memberikan *positional encoding* yang detailnya dapat dilihat di penelitian Vaswani dkk., (2017) pada bagian 3.5.

Terdapat 3 mekanisme atensi yang ada dalam transformer. *Self-attention* pada *encoder* yaitu mekanisme untuk mengaitkan posisi yang berbeda dari suatu sekuens untuk menghitung representasi dari sekuens tersebut. Kemudian *encoder-decoder attention* yaitu mekanisme atensi yang digunakan agar bisa menangkap penekanan pada konteks dari *encoder* pada layer *decoder*. Terakhir *masked self-attention* pada decoder agar model hanya dapat mengaitkan posisi yang berbeda hubungan antar masukan sampai pada urutan masukan tersebut.

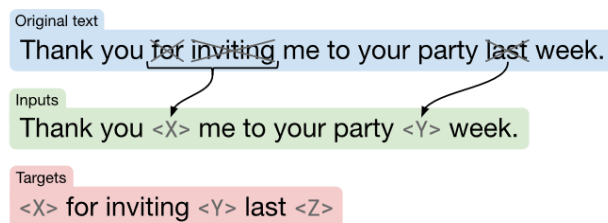
Arsitektur model transformer dapat dilihat pada Gambar II.4. Transformer sering digunakan sebagai dasar arsitektur model bahasa pralatih seperti pada model bahasa pralatih T5 dan BART.



Gambar II.4 Arsitektur Transformer (Vaswani, dkk., 2017)

Salah satu model bahasa pralatih berbasis *encoder-decoder* transformer adalah *Text to Text Transfer Transformers* atau disingkat T5. T5 diajukan Raffel dkk., (2020) sebagai salah satu model pralatih yang memanfaatkan pembelajaran transfer menggunakan pendekatan *text-to-text* untuk memodelkan persoalan pemrosesan bahasa dengan cara menerima suatu teks sebagai masukan dan memberikan keluaran teks (Raffel, dkk., 2020).

Proses pelatihan tanpa supervisi model T5 dilakukan dengan menggunakan objektif “*denoising objective*” atau “*masked language modelling*”. Artinya, model dilatih untuk memprediksi kata-kata yang dihilangkan dari suatu teks (Raffel, dkk., 2020). Ilustrasi objektif dari proses pelatihan tersebut dapat dilihat pada Gambar II.5. Proses pelatihan tersebut dilakukan dengan menggunakan korpus Colossal Clean Crawled Corpus (C4) untuk model T5. Korpus C4 untuk bahasa Inggris berjumlah 364,868,901 teks sebagai data latih dan 364,608 teks sebagai data validasi.



Gambar II.5 Ilustrasi Objektif Pelatihan Tanpa Supervisi dari Proses Pre-training T5 (Raffel, dkk., 2020)

Pada tahap *fine-tuning*, dilakukan perbaruan terhadap semua parameter model T5. Masukan dan keluaran model perlu diubah menjadi teks agar sesuai dengan pendekatan *text-to-text* yang digunakan T5. Proses pelatihan dan inferensi dari T5 agar bisa membangkitkan teks sebagai keluaran model dapat dilihat pada **Lampiran B**.

Model bahasa pralatih T5 untuk bahasa Indonesia yang tersedia adalah IndoT5 (Wikidepia, 2021). Model tersebut dilatih pada korpus berbahasa Indonesia yang diambil dari korpus mC4 (Xue, dkk., 2021) yang sudah difilter. Proses pelatihan model tersebut mengikuti tahapan-tahapan pelatihan pada model T5-base versi 1.1 (https://huggingface.co/google/t5-v1_1-base).

Model dilatih dengan korpus bahasa Indonesia yang diambil dari mC4 (Xue, dkk., 2021) yang telah dipraproses dengan tahapan pelatihan mengikuti model T5-base versi 1.1. Korpus mC4 bahasa Indonesia sendiri terdiri dari 69.625.551 teks data latih dan 69739 teks data validasi. Konfigurasi model bahasa pralatih IndoT5 dapat dilihat pada Tabel II.1.

Tabel II.1 Konfigurasi Model Bahasa Pralatih IndoT5

No	Konfigurasi	Nilai
1	<i>Vocabulary size</i>	32128
2	<i>Number of encoder</i>	12
3	<i>Number of decoder</i>	12
4	<i>Embedding Vector Size</i>	768
5	<i>Number of Heads (Multi-head attention)</i>	12
6	<i>FFNN Params (per blok)</i>	2048
7	<i>Dropout Rate</i>	0.1

Selain T5, model bahasa pralatih berbasis transformer lain adalah BART. BART adalah suatu *denoising autoencoder* untuk model pralatih berbasis *sequence-to-sequence* seperti T5. Perbedaan utama BART dengan T5 terletak pada objektif pralatih dimana model BART melakukan *text infilling* (Menghilangkan sejumlah

rentang kalimat dengan token [MASK]) diikuti *sentence permutation* (mengacak urutan kata dalam teks) dengan tujuan agar model dapat memprediksi token yang dihilangkan serta dapat merekonstruksi teks kembali ke teks asli. Penelitian terkait model bahasa pralatih BART untuk bahasa Indonesia sudah dilakukan oleh Cahyawijaya dkk. (2021) dan menghasilkan model bernama IndoBART. Implementasi model tersebut dibuat berdasarkan model mBART dan memiliki jumlah parameter sekitar 132 juta parameter.

II.6 Evaluasi Kinerja Ekstraksi Triplet Opini

Untuk mengukur kinerja model, tugas akhir menggunakan metrik yang digunakan dalam persoalan klasifikasi yaitu *F1-score*. *F1-score* adalah salah satu indikator kinerja model pembelajaran mesin untuk menilai tingkat akurasi model pada klasifikasi biner. Metrik ini sering disebut sebagai rata-rata harmonik dari *precision* dan *recall* model karena *F1-score* dapat diperoleh dengan menggabungkan *precision* dan *recall*. Persamaan untuk mendapatkan nilai *F1-score* dapat mengacu pada persamaan II-2 (Olson & Delen, 2008).

$$F1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = 2 \frac{Precision * Recall}{Precision + Recall} \quad (II-2)$$

Sementara itu, untuk mendapatkan nilai *precision* dan *recall* dapat mengacu pada persamaan II-3 dan II-4 (Olson & Delen, 2008).

$$Precision = \frac{TP}{TP + FP} \quad (II-3) \quad Recall = \frac{TP}{TP + FN} \quad (II-4)$$

Untuk mengukur kinerja model pada *task* ekstraksi triplet opini, maka teknik evaluasi yang digunakan adalah *exact match*, artinya keluaran dari model ekstraksi triplet opini harus tepat dan hasil prediksi yang bersifat benar sebagian akan dianggap gagal untuk diekstraksi atau salah. Dengan demikian, *true positive* (TP) berarti setiap prediksi triplet opini benar yang terdapat pada label data awal, *false positive* (FP) berarti setiap prediksi triplet opini yang dibangkitkan model namun bukan termasuk triplet opini yang didefinisikan dari label data, dan *false negative*

(FN) berarti setiap triplet opini awal yang tidak berhasil diprediksi atau dibangkitkan oleh model.

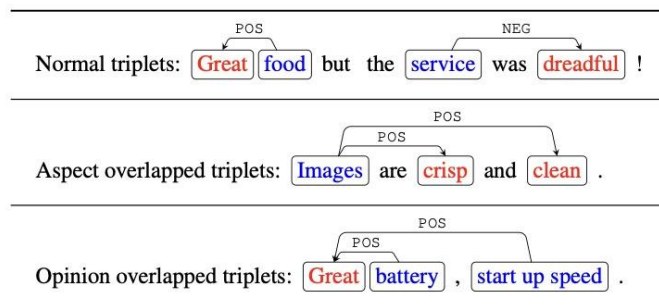
Sebagai contoh, apabila terdapat 100 triplet opini yang akan dievaluasi dan model juga menghasilkan 100 triplet opini yang memiliki 70 triplet opini benar, 30 triplet opini yang tidak terdapat pada label awal, dan 30 triplet opini gagal diekstraksi model, maka nilai *F1-score* untuk kasus tersebut dapat dihitung sebagai berikut.

$$F1 = 2 \frac{Precision * Recall}{Precision + Recall} = 2 \frac{0.7 * 0.7}{0.7 + 0.7} = 0.7$$

II.7 Penelitian Terkait

II.7.1 Penelitian Wirawan (2021)

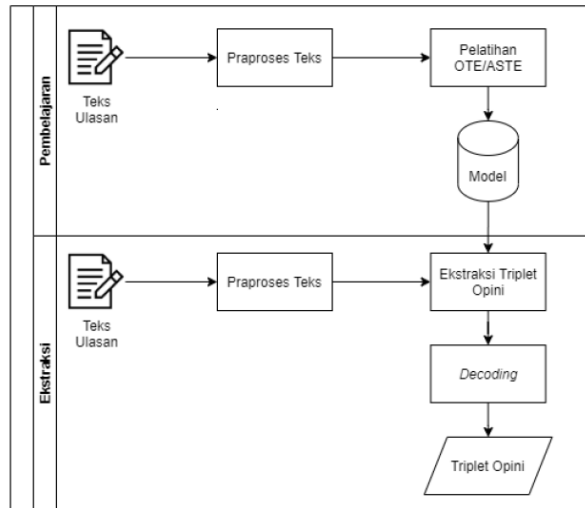
Penelitian Wirawan (2021) berfokus pada penyelesaian ekstraksi triplet opini menggunakan pendekatan *grid tagging scheme* (GTS) (Wu, dkk., 2020). Pendekatan tersebut melakukan transformasi *task* ekstraksi pasangan opini menjadi sebuah kesatuan pelabelan grid yang dimodifikasi untuk melakukan ekstraksi triplet opini.



Gambar II.6 Ilustrasi Jenis Triplet (Zhang, dkk., 2020)

Dalam ekstraksi triplet opini dapat terjadi fenomena tumpang tindih (*overlapped triplet*) yaitu dua buah triplet terdiri dari ekspresi aspek/sentimen yang sama yang dapat dilihat pada Gambar II.6 karena satu ekspresi aspek dapat dirujuk oleh beberapa ekspresi sentimen dan juga sebaliknya. Tugas akhir Wirawan (2021) bertujuan menyelesaikan permasalahan triplet opini yang tumpang tindih.

Untuk menyelesaikan permasalahan tersebut, Wirawan memanfaatkan 4 (empat) buah modul yaitu modul praproses teks, modul pelatihan model ekstraksi triplet opini, modul ekstraksi triplet opini, dan modul *decoding*. Sebagai gambaran, arsitektur yang digunakan pada tugas akhir Wirawan (2021) dapat dilihat pada Gambar II.7.



Gambar II.7 Arsitektur Sistem Ekstraksi Triplet Opini pada Tugas Akhir Wirawan (2021)

Model tersebut dilatih pada kumpulan ulasan hotel dengan sedikit modifikasi pada anotasi relasi triplet opini. Pada penelitian ini juga dikenalkan istilah kalimat triple dan kalimat campuran, dimana kalimat triple artinya setiap ekspresi aspek dan ekspresi sentimen memiliki pasangan masing-masing, sedangkan kalimat campuran berarti kalimat memiliki satu atau lebih ekspresi yang tidak memiliki pasangan. Data eksperimen kemudian dibagi menjadi 2 (dua) yaitu data latih *uncleaned* (berisi kalimat campuran) dan data latih *cleaned* (tidak berisi kalimat campuran). Hasil evaluasi *F1-score* yang didapat dari task ekstraksi triplet opini dengan konfigurasi terbaik adalah 0.784. Detail hasil evaluasi dapat dilihat pada Tabel II.2.

Tabel II.2 Hasil Kinerja Terbaik Model Tugas Akhir Wirawan (2021)

Model	Ekstraksi Triplet Opini		
	Precision	Recall	F1-Score
TA	0.8096	0.7600	0.7840

Berdasarkan hasil analisa, model masih memiliki beberapa kelemahan seperti kesulitan apabila melakukan ekstraksi ekspresi aspek dan sentimen pada token yang belum pernah ada atau jarang muncul pada data latih. Selain itu, model masih kurang baik dalam mengekstraksi pasangan triplet opini yang memiliki ekspresi sentimen di depan ekspresi aspek meskipun model telah berhasil mengekstraksi dengan baik ekspresi aspek dan ekspresi sentimen dalam teks ulasan. Contoh kegagalan tersebut dapat dilihat pada Tabel II.3.

Tabel II.3 Contoh Kegagalan Ekstraksi Triplet Opini pada Sentimen yang Terletak di Depan Aspek (Wirawan, 2021)

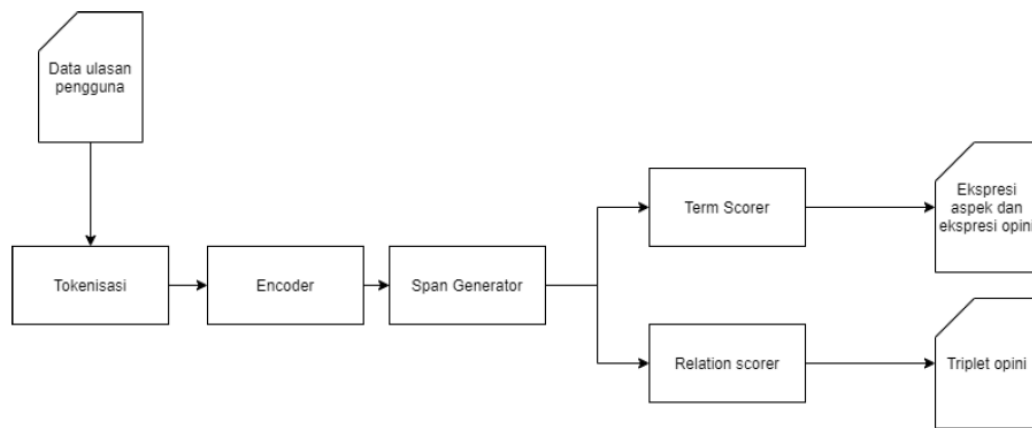
Kalimat	bagus tempatnya , bersih dan nyaman .
Triplet Sebenarnya	(['tempatnya'], ['bagus'], 'PO'), (['tempatnya'], ['bersih'], 'PO'), (['tempatnya'], ['nyaman'], 'PO')
Triplet Hasil Prediksi	- (X) (['tempatnya'], ['bersih'], 'PO'), (['tempatnya'], ['nyaman'], 'PO')

II.7.2 Penelitian Genadi (2021)

Penelitian Genadi (2021) berfokus pada penyelesaian permasalahan ekstraksi triplet opini dengan memanfaatkan representasi *span* yang akan digunakan pada *framework* SpanMLT (Zhao, dkk., 2020). Penyesuaian yang dilakukan adalah pada bagian *relation scorer* sehingga *relation scorer* tidak hanya melakukan klasifikasi biner ada tidaknya relasi pada suatu pasangan *span*, tetapi melakukan klasifikasi multikelas apakah relasi *span* tersebut positif, negatif, atau tidak berelasi.

Tesis Genadi bertujuan untuk menyelesaikan permasalahan utama penggunaan pendekatan *sequence labelling* untuk ekstraksi triplet opini yaitu inkonsistensi hasil prediksi polaritas sentimen pada satu ekspresi. Hal ini terjadi karena klasifikasi polaritas sentimen pada *sequence labeling* dengan *joint-tagging* dilakukan dengan cara memberikan label polaritas $P \in \{PO, NG, O\}$ pada setiap token yang ada dalam kalimat. Artinya, terdapat kemungkinan adanya perbedaan prediksi polaritas pada token yang seharusnya tergabung dalam satu ekspresi (Hu, dkk., 2019).

Tesis Genadi (2021) membentuk perancangan solusi serupa dengan *framework* SpanMLT dari penelitian Zhao dkk., (2020) dengan modifikasi *relation scorer*. Sebagai gambaran, arsitektur yang digunakan dalam penelitian Genadi (2021) dapat dilihat pada Gambar II.8.



Gambar II.8 Arsitektur Sistem Ekstraksi Triplet Opini Berbasis Span pada Tesis Genadi (2021)

Model tersebut dijalankan pada kumpulan ulasan hotel yang disediakan oleh pihak AiryRooms. Data eksperimen terdiri dari 5000 teks ulasan dengan jumlah token mencapai 78603 token, yang memiliki panjang rata-rata per kalimat sebesar 15.72 kata. Angka akurasi yang diperoleh pada model tesis untuk ekstraksi triplet opini (ASTE) sebesar 0.56, berbeda jauh dengan model TA Genadi yaitu 0.74.

Tesis Genadi (2021) mengutarakan kelemahan model seperti banyaknya relasi yang tidak berhasil diekstraksi oleh model tesis. Saat *framework* SpanMLT diujicobakan, banyak kandidat *span* yang seharusnya ada di dalam relasi malah tidak tergabung

ke kandidat *span* yang dipilih (berdasarkan peluang) bahkan pada data latih sendiri. Hal ini terjadi karena tidak semua kandidat *span* diambil untuk proses inferensi model mengingat adanya keterbatasan infrastruktur sehingga tidak memungkinkan apabila seluruh kandidat *span* yang dibangkitkan diproses.

Penanggulangan untuk permasalahan tersebut adalah dengan penambahan *relation scorer* untuk mengukur probabilitas ada tidaknya *span* di dalam suatu relasi. Namun, hasil analisa menunjukkan penanganan tersebut masih belum cukup untuk memperbaiki kinerja model tesis dalam melakukan ekstraksi triplet opini. Selain itu, perlunya pembatasan pencarian persentase jumlah kandidat *span* yang dipilih serta besar *span* akibat penggunaan memori yang terlalu besar juga membatasi peningkatan kinerja model saat pelatihan.

II.7.3 Penelitian Rendyanto (2021)

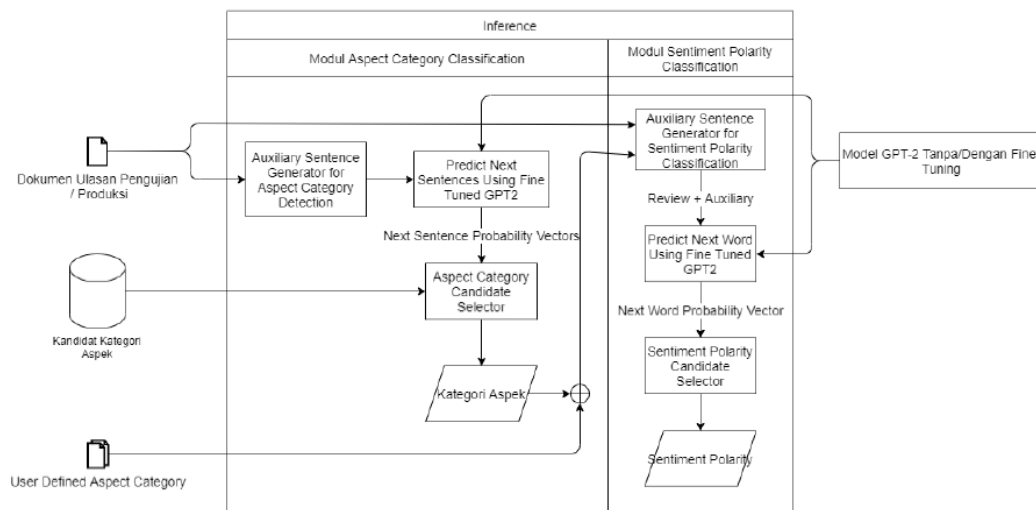
Penelitian Rendyanto (2021) berfokus pada pendekatan pembangkitan teks dengan model bahasa pralatih. Hasil pembelajaran kemudian dimanfaatkan untuk menyelesaikan *task* klasifikasi polaritas sentimen dan klasifikasi jenis aspek dari teks ulasan. Penelitian ini terinspirasi dari penelitian di bidang AI yang menunjukkan bahwa model *deep learning* yang dilatih dengan metode pembelajaran tanpa supervisi, seperti model bahasa *Generative Pre-Trained Transformer 2* (GPT-2) mampu mencapai kinerja *state of the art* pada berbagai benchmark *task* pemodelan bahasa.

Alasan penelitian ini dilakukan oleh Rendyanto adalah penggunaan model GPT-2 dapat menyelesaikan banyak permasalahan di bidang pemrosesan bahasa alami, seperti dalam pemahaman membaca, penerjemahan mesin, menjawab pertanyaan dan peringkasan. Namun sayangnya, penelitian terkait model generatif di bidang analisis sentimen berbasis aspek masih jarang dilakukan. Untuk itu, dengan menggunakan penelitian Brown dkk., (2020) mengenai model generatif sebagai inspirasi, Rendyanto ingin menyelesaikan permasalahan *out of vocabulary* (OOV) dan permasalahan kurangnya dataset berlabel dengan model bahasa pralatih yang

diuji secara *cross-domain* untuk mengetahui kinerja model yang dihasilkan pada domain umum.

Untuk menyelesaikan *task* klasifikasi polaritas sentimen dan kategorisasi aspek tersebut, Rendyanto memanfaatkan empat buah modul, yaitu modul klasifikasi polaritas sentimen, modul deteksi kategorisasi aspek, modul *fine tuning* dan modul *oversampling*. Sebagai gambaran, arsitektur yang digunakan dalam penelitian Rendyanto (2021) dapat dilihat pada Gambar II.9.

Ide solusi utama dari penelitian tugas akhir Rendyanto (2021) adalah dengan menambahkan *auxiliary sentence* pada akhir dokumen ulasan untuk mengubah permasalahan analisis sentimen berbasis aspek menjadi permasalahan pembangkitan teks yang merupakan permasalahan yang digunakan untuk melatih model GPT-2. *Auxiliary sentence* yang ditambahkan berupa kalimat pertanyaan yang bertujuan sebagai pemicu agar model membangkitkan teks berupa jawaban dari pertanyaan tersebut yang digunakan sebagai keluaran model. *Auxiliary sentence* yang digunakan disesuaikan dengan *task* yang diselesaikan.



Gambar II.9 Arsitektur Sistem Klasifikasi Kategori Aspek dan Klasifikasi Polaritas Sentimen Tugas Akhir Rendyanto (2021)

Model tersebut dijalankan pada dataset domain restaurant SemEval 2016 Task 5, dataset domain laptop SemEval 2016 Task 5, dan dataset domain restaurant SemEval 2014 Task 4. yang tersedia secara publik oleh SemEval. Hasil evaluasi pada konfigurasi terbaik hasil eksperimen dapat dilihat pada Tabel II.4.

Berdasarkan hasil tersebut, Rendyanto menyimpulkan bahwa model memiliki kapabilitas yang baik dalam permasalahan klasifikasi polaritas sentimen. Apabila model perlu digunakan untuk domain lainnya, model tidak perlu dilatih ulang karena telah bersifat general secara umum. Namun, pada permasalahan pendeteksian kategori aspek sangat bergantung pada proses *fine tuning* yang dilakukan sehingga masih diperlukan *fine tuning* model pada domain lain.

Tabel II.4 Hasil Kinerja Terbaik Model Tugas Akhir Rendyanto (2021)

Dataset	Metrik Evaluasi	Permasalahan	
		Klasifikasi Polaritas Sentimen	Deteksi Kategori Aspek
Restaurant SemEval 2016 Task 5 Sub Task 1	F1	0.912	0.672
	Akurasi	92.5	-
Laptop SemEval 2016 Task 5 Sub Task 1	F1	0.909	0.509
	Akurasi	92.0	-
SemEval 2016 Task 5 Sub Task 1 (Restaurant + Laptop)	F1	0.910	-
	Akurasi	92.2	-
Restaurant SemEval 2014 Task 4	F1	0.687	0.825
	Akurasi	79.3	-

BAB III

ANALISIS MASALAH DAN PERANCANGAN SOLUSI

III.1 Analisis Persoalan

Untuk mengadaptasi ekstraksi triplet opini dengan pendekatan pembangkitan teks yang diusulkan Zhang dkk., (2021) diperlukan *resource*, yaitu:

1. Korpus bahasa Indonesia untuk melakukan ekstraksi triplet opini pada proses eksperimen *fine-tuning* dan pengujian model. Korpus yang digunakan pada penelitian Zhang dkk., (2021) terdiri dari 1075-2068 teks ulasan. Sementara itu, korpus teks ulasan bahasa Indonesia yang digunakan pada tugas akhir ini adalah data ulasan pada domain perhotelan yang digunakan pada tugas akhir Wirawan (2021) yang terdiri dari 5000 kalimat ulasan. Namun, data yang digunakan pada tugas akhir Wirawan (2021) ditujukan untuk ekstraksi triplet opini dengan pendekatan *sequence labelling* dengan melabeli tiap token dengan sebuah tag. Agar dapat digunakan pada pendekatan pembangkitan teks, diperlukan perubahan pada format data seperti format *extraction-style* atau *annotation-style* seperti pada penelitian Zhang dkk., (2021) yang dapat digunakan untuk mendapatkan triplet opini setelah pembangkitan teks berhasil dilakukan.
2. Model bahasa pralatih yang mendukung bahasa Indonesia. Pada penelitian Zhang dkk. (2021), model bahasa pralatih yang digunakan adalah T5-base untuk bahasa Inggris yang memiliki kosa kata yang berbeda dari bahasa Indonesia.
3. Langkah *postprocess* yang cocok untuk bahasa Indonesia. Pada penelitian Zhang dkk., (2021), langkah *postprocess* berupa normalisasi prediksi dilakukan pada token hasil pembangkitan yang berbeda dari teks masukan untuk memperbaiki perubahan dari kata tunggal menjadi jamak ataupun sebaliknya pada bahasa Inggris. Fenomena tersebut tidak terjadi pada bahasa Indonesia. Selain itu, strategi normalisasi yang digunakan umumnya

gagal pada kasus ketika model membangkitkan kata yang cukup berbeda secara leksikal dari target (Zhang dkk., 2021).

Beberapa triplet opini menunjukkan polaritas sentimen yang kontradiktif terhadap suatu aspek. Sebagai contoh, pada teks ulasan “kamar bersih , kamar mandi bersih , kolam renang baik , classic atmosphere” salah satu triplet opininya yaitu (kamar, bersih, **negatif**) memiliki polaritas negatif padahal polaritas seharusnya adalah positif. Selain itu ditemukan juga kata dalam kalimat ulasan dengan campuran bahasa (*code-mixing*), contoh kata dengan kasus tersebut seperti kata “*recommended*”, “*good*”, dan “*overall*” yang sering muncul pada kalimat ulasan (Wirawan, 2021).

Analisis data yang dilakukan terdapat dataset yang digunakan pada tugas akhir dengan dataset penelitian Zhang dkk., (2021) menunjukkan perbedaan yaitu terdapat kalimat pernyataan ataupun saran yang tidak mengandung ekspresi sentimen. Contoh dari kasus ini dapat dilihat pada kalimat pada data latih “sudah sering menginap disini”. Kalimat tersebut tidak memiliki opini yang ingin diutarakan dalam kalimat ulasan. Kasus ini terjadi pada 261 dari 5000 kalimat pada dataset. Pada tugas akhir, kasus ini tidak ditangani secara khusus, namun dianalisis kemampuan model tugas akhir dalam menangani kasus ini.

Triplet opini dengan aspek implisit dapat ditemukan pada ekspresi sentimen yang tidak memiliki pasangan ekspresi aspek. Saat ini, dataset yang tersedia masih ditujukan untuk ekstraksi triplet opini eksplisit. Contoh dari kasus ini ada pada kalimat ulasan “**bagus** dan **bersih**, namun pelayan **cerewet**.”. Pada kalimat tersebut ekspresi sentimen “**bagus**” dan “**bersih**” merujuk pada aspek yang tidak dituliskan eksplisit dalam ulasan. Agar ekspresi sentimen tersebut dapat dipasangkan membentuk pasangan triplet opini dengan aspek implisit, diperlukan metode anotasi untuk aspek implisit serta anotasi tambahan untuk mendapatkan pasangan triplet opini dengan aspek implisit.

III.2 Analisis Solusi

Perbaikan kualitas korpus pada kasus triplet opini dengan polaritas sentimen yang kontradiktif dilakukan dengan perbaikan anotasi triplet opini tersebut. Misal, pada ulasan "kamar mandi bau banget...", salah satu triplet opini ulasan tersebut adalah (kamar mandi, bau banget, positif) padahal polaritas seharusnya adalah negatif. Keseluruhan perbaikan polaritas triplet opini dapat dilihat pada **Lampiran A.1**. Sementara itu, kalimat ulasan dengan campuran bahasa (*code-mixing*) dan kalimat pernyataan ataupun saran yang tidak mengandung ekspresi sentimen tidak dilakukan penanganan khusus pada kasus tersebut agar dapat ditangani secara langsung oleh model bahasa pralatih yang digunakan dalam tugas akhir.

III.2.1 Konstruksi Korpus dengan Format *Extraction-style*

Agar dapat mendukung pendekatan pembangkitan teks, dilakukan konstruksi ulang korpus yang mengambil inspirasi dari format *extraction-style* pada penelitian Zhang dkk., (2021). Format *extraction-style* menyimpan informasi terkait letak komponen triplet opini pada teks ulasan untuk membentuk data latih. Visualisasi anotasi data dan target dengan format ini dapat dilihat pada Tabel III.1. Pada kasus ini dilakukan konstruksi data latih pada kalimat ulasan “kamarnya bagus, tetapi toilet kurang bersih”.

Tabel III.1. Konstruksi Korpus dengan Format *Extraction-style*

Teks Ulasan	kamarnya bagus, tetapi toilet kurang bersih.
Anotasi Data	kamarnya bagus, tetapi toilet kurang bersih. #####([([0],[1],”POS”),([4],[5,6], “NEG”)])
Target	(kamarnya, bagus, positif); (toilet, kurang bersih, negatif)

Alternatif lain dalam konstruksi korpus dapat menggunakan format *annotation-style* seperti yang telah dijelaskan pada subbab II.2. Konstruksi korpus dengan format *extraction-style* dipilih karena menghasilkan keluaran yang lebih pendek dibanding format *annotation-style* dan terbukti lebih baik digunakan ketimbang

annotation-style pada ekstraksi triplet opini berdasarkan penelitian Zhang dkk., (2021).

III.2.2 Model Bahasa Pralatih

Model bahasa pralatih yang dipilih dalam penelitian tugas akhir ini adalah IndoT5 yang dilatih dengan data C4 untuk bahasa Indonesia sebesar 242 GB. Model ini dipilih sebagai bentuk penyesuaian terhadap model bahasa pralatih yang digunakan pada teknik penyelesaian masalah penelitian Zhang dkk. (2021) untuk bahasa Indonesia. Alternatif model bahasa pralatih lainnya yang dapat digunakan seperti IndoBART (Cahyawijaya dkk., 2021) yang dilatih pada korpus Indo4B.

Pemilihan model bahasa pralatih dilakukan berdasarkan eksplorasi awal model bahasa pralatih untuk ekstraksi triplet opini dengan pendekatan pembangkitan teks, dimana model IndoBART mengalami kesulitan untuk menghasilkan teks dengan format *extraction-style*. Hampir seluruh teks hasil bangkitan model memiliki format yang tidak teratur sehingga tidak diabaikan. Contoh perbandingan keluaran yang dihasilkan dari teks ulasan masukan kedua model dapat dilihat pada Tabel III.2.

Tabel III.2 Contoh Hasil Inferensi Model Bahasa Pralatih

Teks Ulasan	“tidak ada snack airy .”
Hasil Inferensi IndoBART	“snack airy, tidak ada, negatif), negatif”
Hasil Inferensi IndoT5	“(snack airy, tidak ada, negatif);”

Penggunaan model bahasa pralatih diharap dapat meningkatkan kinerja model dan mencegah kasus OOV, yaitu dengan teknik tokenisasi *sentencepiece* yang dilakukan T5. Selain itu, penelitian ini juga mengujikan efektifitas *pre-training* lanjutan (*posttraining*) sebagai variasi model bahasa yang digunakan dengan menggunakan data korpus domain hotel berjumlah 136.788 kalimat (9.4 MB) dari penelitian Aditya (2021) agar model pralatih dapat mempelajari terminologi domain hotel secara umum.

III.2.3 Langkah *Postprocess*

Seperti yang telah dijelaskan pada subbab II.2, teknik penyelesaian masalah yang diajukan pada penelitian Zhang dkk., (2021) memiliki suatu langkah *postprocess* setelah proses pembangkitan triplet opini berhasil dilakukan. Langkah tersebut berupa pengecekan apakah kosa kata ekspresi aspek dan sentimen dari triplet opini hasil bangkitan sesuai dengan teks masukan atau tidak. Sebagai contoh, pada ulasan “hotelnya bersih” jika model membangkitkan triplet opini (hotel, bersih, positif), maka aspek dalam triplet opini tidak terdapat pada teks ulasan. Langkah *postprocess* bertujuan untuk mengecek dan memperbaiki kasus tersebut.

Hal ini dilakukan karena idealnya setiap aspek dan sentimen hasil bangkitan model seharusnya tepat sesuai dengan kumpulan kosa kata dari teks masukan kecuali untuk aspek implisit (NULL) pada sentimen tanpa pasangan untuk triplet opini dengan aspek implisit. Namun, hal ini tidak selalu berlaku karena setiap keluaran yang dibangkitkan berasal dari kumpulan kosa kata model dan bukan kumpulan kosa kata dari teks masukan.

Untuk itu, langkah *postprocess* diterapkan untuk setiap keluaran triplet opini hasil bangkitan model yang tidak terdapat pada kumpulan kosa kata teks masukan awal. Hal ini dilakukan dengan mencari kosa kata paling mirip pada kumpulan kosa kata teks masukan dari kosa kata yang bermasalah. Teknik perhitungan jarak antar teks yang digunakan pada tugas akhir ini adalah dengan menggunakan jarak levenshtein seperti yang diusulkan pada penelitian Zhang dkk., (2021) dan *cosine similarity* menggunakan vektor yang dihasilkan dari *word embedding* seperti pada penelitian Sitikhu dkk., (2019).

Cosine similarity bertujuan untuk menangani kasus kesalahan ketika model membangkitkan kata yang cukup berbeda secara leksikal dari target namun masih memiliki makna yang mirip. *Word embedding* yang digunakan untuk mendapatkan vektor representasi kosa kata adalah layer *embedding* pada model bahasa pralatih IndoT5.

III.2.4 Anotasi Triplet Opini dengan Aspek Implisit

Untuk melengkapi dataset dengan pasangan triplet opini dengan aspek implisit dilakukan anotasi tambahan pada ekspresi sentimen yang tidak memiliki pasangan ekspresi aspek untuk membentuk pasangan triplet opini dengan aspek implisit. Anotasi tambahan dilakukan dengan menambahkan anotasi polaritas sentimen berdasarkan ekspresi sentimen. Karena aspek pada ekspresi sentimen yang tidak memiliki pasangan ekspresi aspek tidak dinyatakan secara eksplisit, aspek ditetapkan dengan simbol karakter “_” untuk menandakan aspek implisit. Contoh dari anotasi yang telah tersedia dapat dilihat pada Tabel III.3 dengan (+) menunjukkan anotasi yang ditambahkan pada tugas akhir ini.

Tabel III.3 Contoh Anotasi Data

Teks Ulasan	
oke banget , tetapi ac nya tidak bisa diatur suhu nya .	
Ekspresi Aspek	Ekspresi Sentimen
- ac nya	- oke banget - tidak bisa diatur
Triplet Opini	(+) Triplet Opini Tanpa Pasangan
- (ac nya, tidak bisa diatur, negatif)	- (_, oke banget, positif)

Langkah anotasi untuk aspek implisit yang dilakukan tugas akhir mengikuti langkah anotasi pada penelitian Cai dkk., (2021) dan Zhang dkk., (2022) yang menganotasi aspek spesial NULL untuk merujuk pada aspek implisit. Dalam hal ini, karakter “_” digunakan untuk merepresentasikan NULL untuk pasangan triplet opini dengan aspek implisit. Proses anotasi diterapkan pada 1483 kalimat yang mengandung ekspresi sentimen yang tidak memiliki pasangan ekspresi aspek. Anotasi dilakukan dengan memberikan indeks “-1” pada label yang menandakan NULL sebagai ekspresi aspek implisit seperti yang dapat dilihat pada Tabel III.4.

Tabel III.4 Contoh Anotasi Triplet Opini dengan Aspek Implisit

Teks Ulasan	“bagus dan bersih.”
Anotasi Data	“bagus dan bersih.” #####[(-1],[0],”POS”), (-1],[2],”POS”)]
Target	(_, bagus, positif); (_, bersih, positif);

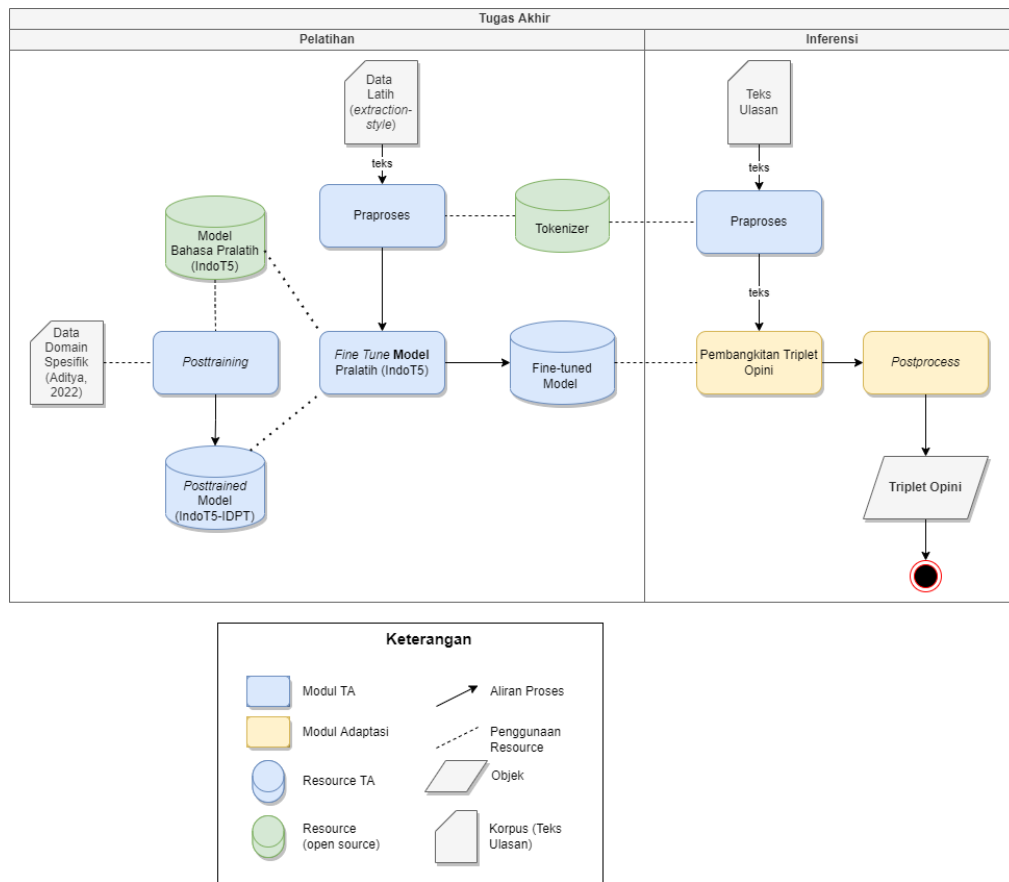
III.3 Rancangan Solusi

Sistem ekstraksi triplet opini yang dapat menangani aspek implisit dengan pendekatan pembangkitan teks pada tugas akhir dibangun dengan mengadaptasi teknik penyelesaian masalah yang diajukan Zhang dkk., (2021). Ide utama solusi tugas akhir adalah modifikasi korpus pada dokumen ulasan sesuai dengan format anotasi *extraction-style* agar ekstraksi triplet opini dapat ditransformasi dari pendekatan *sequence labelling* menjadi pembangkitan teks dengan penggunaan model bahasa pralatih yang dilatih untuk membangkitkan triplet opini dari teks ulasan sesuai dengan format *extraction-style*.

Arsitektur solusi sistem ditunjukkan pada Gambar III.1. Pada gambar tersebut terdapat 5 modul utama untuk membangun sistem, yaitu:

1. Modul Praproses Teks Masukan
2. Modul *Fine-tune* Model Ekstraksi Triplet Opini
3. Modul *Posttraining*
4. Modul Pembangkitan Triplet Opini
5. Modul *Postprocess*

Sebelum proses pelatihan dilakukan, dilakukan konstruksi label dengan format *extraction-style* untuk tiap teks ulasan pada data latih sebagai target dalam proses *fine-tuning* model bahasa pralatih untuk mendukung pendekatan pembangkitan teks seperti yang telah dijelaskan pada subbab III.2.1.



Gambar III.1. Arsitektur Sistem Ekstraksi Triplet Opini dengan Pendekatan Pembangkitan Teks

Selain itu, anotasi tambahan untuk triplet opini dengan aspek implisit dilakukan pada ekspresi sentimen yang tidak memiliki pasangan ekspresi aspek untuk membentuk pasangan triplet opini dengan aspek implisit seperti yang telah dijelaskan pada subbab III.2.4. Anotasi dilakukan dengan memasang ekspresi sentimen tanpa pasangan aspek dengan sebuah aspek spesial (NULL) kemudian dilanjutkan dengan penambahan informasi berupa anotasi polaritas sentimen (positif, negatif, netral) pada pasangan triplet opini dengan aspek implisit tersebut. Triplet opini tambahan tersebut digunakan pada proses *fine-tuning* ekstraksi triplet opini dengan penanganan terhadap aspek implisit. Hasil anotasi tambahan triplet opini dengan aspek implisit dapat dilihat pada **Lampiran A.3**.

III.3.1 Modul Praproses

Modul ini bertujuan untuk melakukan praproses pada teks masukan dan mengubah representasi kalimat masukan yang berupa string menjadi sekuens ID kata yang diperoleh dari kamus kosa kata *tokenizer* model bahasa pralatih.

Tahapan praproses yang diterapkan pada modul ini yaitu:

1. Perubahan huruf kapital menjadi non kapital untuk mengurangi variasi teks
2. Penambahan padding pada bagian belakang dari teks untuk menyamakan besar dimensi masukan model (128) dengan *tokenizer* model.
3. Pemotongan teks jika hasil tokenisasi subkata melebihi batas yang telah ditentukan (128) dengan *tokenizer* model.

Penggunaan *tokenizer* juga melakukan pembersihan spasi, pemisahan tanda baca, serta tokenisasi dan pemetaan ID kata sesuai kamus *tokenizer*. Contoh masukan dan keluaran modul praproses teks dapat dilihat pada Tabel III.5.

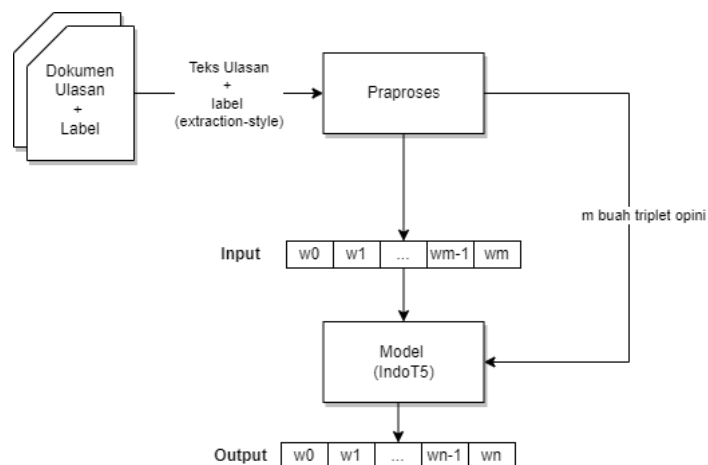
Tabel III.5 Contoh Masukan dan Keluaran Modul Pengolahan Teks

Teks Masukan	Keluaran Modul Praproses	
Kamarnya bagus, tetapi Toilet kurang bersih	<i>lowercase</i>	kamarnya bagus, tetapi toilet kurang bersih.
	<i>tokenisasi</i>	[_kamar, nya, _bagus, _, _tetapi, _toilet, _kurang, _bersih, _., </s>]
	<i>padding</i>	[_kamar, nya, _bagus, _, _tetapi, _toilet, _kurang, _bersih, _., </s>, <pad>,...<pad>]
	<i>input ids</i>	[2792, 23, 3805, 11, 8, 93, 13912, 523, 4113, 11, 3, 1, 0, ..., 0]
	<i>attention mask</i>	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0,...0]

III.3.2 Modul *Fine-Tune* Model Ekstraksi Triplet Opini

Modul *fine-tune* model bahasa pralatih untuk ekstraksi triplet opini adalah modul yang ditujukan untuk melatih model bahasa pralatih untuk ekstraksi triplet opini. Model yang digunakan adalah model bahasa pralatih IndoT5 yang dilatih untuk *task* ekstraksi triplet opini eksplisit serta triplet opini dengan aspek implisit. Visualisasi kerja modul dapat dilihat pada Gambar III.2. Target model adalah 1 kalimat yang

berisi M buah triplet opini. Kalimat yang dibentuk berisi triplet opini dari dokumen ulasan masukan dengan format *extraction-style* yang dipisahkan dengan simbol ";".

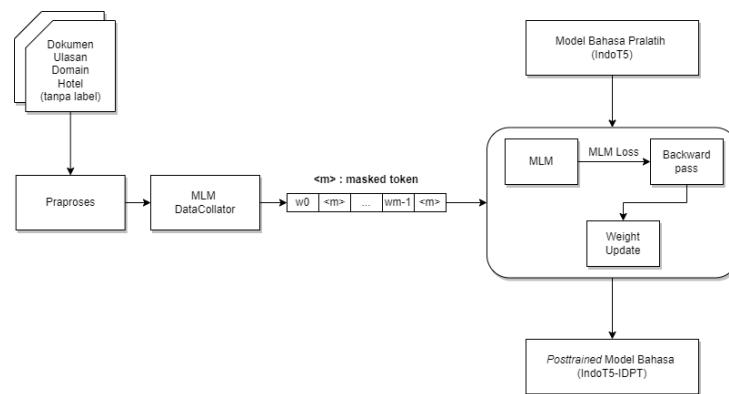


Gambar III.2 Visualisasi Kerja Modul Fine-Tune Model Ekstraksi Triplet Opini Model kemudian dilatih (*fine-tuning*) untuk memprediksi kalimat selanjutnya dari dokumen ulasan yang berisi m buah triplet opini tersebut. Model bahasa pralatih yang dipilih dalam penelitian tugas akhir ini adalah IndoT5 yang dilatih dengan data C4 untuk bahasa Indonesia sebesar 242 GB. Model ini dipilih sebagai bentuk penyesuaian terhadap model bahasa pralatih yang digunakan pada teknik penyelesaian masalah penelitian Zhang dkk., (2021) untuk bahasa Indonesia.

Proses pembelajaran diterapkan agar bagian *encoder* model dapat mempelajari bentuk representasi teks kontekstual yang kemudian menjadi masukan untuk *decoder* model yang mempelajari proses membangkitkan triplet opini dengan format *extraction-style* dari hasil representasi encoder. Dengan teknik pembelajaran ini model dapat melakukan ekstraksi triplet opini dan menghasilkan keluaran triplet opini yang dapat diekstraksi berdasarkan format *extraction-style*. Kinerja model kemudian dapat dievaluasi dengan metrik *F1-score* seperti yang telah dijelaskan pada subbab II.6.

III.3.3 Modul *Posttraining*

Pada pemrosesan bahasa alami, *post-training* adalah proses pelatihan lanjutan terhadap model bahasa pralatih seperti BERT, GPT, T5, dan lain-lain. *Post-training* melanjutkan pelatihan tanpa supervisi model pada data tidak berlabel untuk meningkatkan pengetahuan model pada domain tertentu (Aditya, 2022). *Post-training* pada tugas akhir ini dilakukan pada model IndoT5 dengan objektif pralatih *masked language modelling* (MLM) sesuai dengan yang telah dijelaskan pada subbab II.5. MLM bertujuan untuk memprediksi kata yang telah di-*mask* dari suatu teks. Hal ini dilakukan untuk mendorong model mempelajari representasi kata opini seperti "enak" atau "buruk" pada token yang telah di-*mask* tersebut (Aditya, 2022). Ilustrasi dari proses *post-training* dapat dilihat pada Gambar III.3.

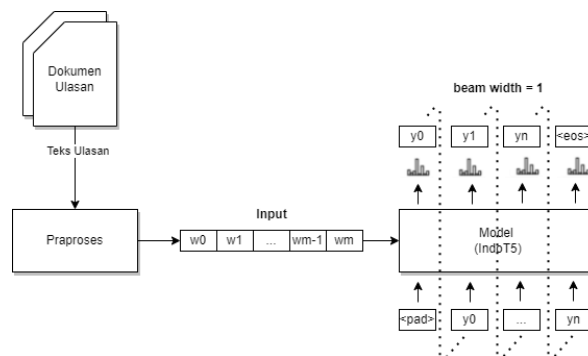


Gambar III.3 Ilustrasi Proses *Post-training* untuk IndoT5

Post-training diterapkan pada model bahasa pralatih IndoT5 sebagai variasi model bahasa yang digunakan dalam tugas akhir. Aditya (2022) menunjukkan penerapan *post-training* meningkatkan kinerja model bahasa pralatih dalam analisis sentimen berbasis aspek. Hal ini dilakukan dengan menggunakan data korpus domain hotel berjumlah 136.788 kalimat (9.4 MB) dari penelitian Aditya (2022) agar model pralatih dapat mempelajari representasi teks terminologi domain hotel secara umum. Hasil dari proses *post-training* pada model IndoT5 adalah *post-trained* IndoT5 yang akan disebut IndoT5-IDPT pada tugas akhir.

III.3.4 Modul Pembangkitan Triplet Opini

Modul pembangkitan triplet opini bertujuan untuk membangkitkan N triplet opini dengan format *extraction-style* untuk setiap dokumen ulasan dimana N adalah banyaknya triplet opini yang berhasil diekstrak model bahasa. Untuk setiap dokumen ulasan yang telah dipraproses, modul menjalankan proses inferensi model bahasa sampai model membangkitkan token akhir kalimat (<eos>) atau mencapai batas maksimal panjang vektor keluaran model. Ilustrasi dari proses inferensi model dapat dilihat pada Gambar III.4.



Gambar III.4 Ilustrasi Proses Pembangkitan Triplet Opini

Model membangkitkan token selanjutnya dengan menentukan persebaran probabilitas kemunculan kata berikutnya untuk semua kata yang ada pada kamus kosa katanya, persebaran ini selanjutnya disebut dengan vektor probabilitas. Vektor probabilitas kemudian dapat di-*decode* oleh tokenizer digunakan untuk mendapatkan triplet opini dari teks ulasan. Pencarian keluaran triplet opini dilakukan dengan algoritma pencarian *beam*.

Keluaran dari proses inferensi model adalah sebuah matriks berukuran *vocab* x panjang sekuens. Ukuran *vocab* adalah besar kamus yang dimiliki IndoT5 yaitu sebesar 32108 kosa kata. Panjang sekuens adalah banyaknya token hasil tokenisasi dari tokenizer dari dokumen ulasan yang telah dipraproses dengan panjang sekuens maksimum sebesar 128. Modul kemudian dapat mengekstrak triplet opini hasil keluaran model sesuai dengan format *extraction-style*. Apabila keluaran model tidak sesuai dengan yang telah ditetapkan sehingga elemen-elemen dalam triplet opini tidak dapat diekstraksi, maka

hasil prediksi diabaikan (dianggap salah dalam evaluasi). Contoh masukan dan keluaran modul ekstraksi triplet opini dapat dilihat pada Tabel III.6.

Tabel III.6 Contoh Masukan dan Keluaran Modul Pembangkitan Triplet Opini

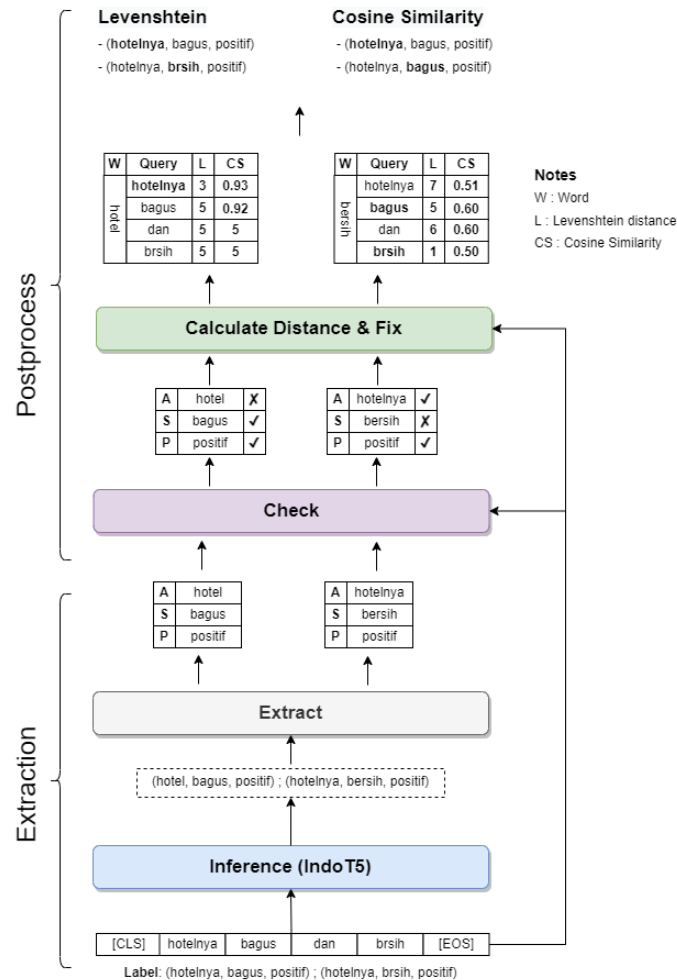
Masukan		Keluaran Modul Pembangkitan Triplet Opini		
Teks Praproses: [“kamarnya”, “bagus”, “,”, “tetapi”, “toilet”, “kurang”, “bersih”, “.”]	Hasil	Hasil Inferensi	“(kamarnya, bagus, positif); (toilet, kurang bersih, negatif)”	
		Hasil Ekstraksi (keterangan) A : Aspek S : Sentimen P : Polaritas	Triplet Opini 1: A: kamarnya S: bagus P: positif	Triplet Opini 2: A: toilet S: kurang bersih P: negatif

III.3.5 Modul *Postprocess*

Modul *postprocess* digunakan untuk mengecek dan memperbaiki hasil pembangkitan triplet yang tidak sesuai dengan kumpulan kosa kata pada teks masukan. Modul menerima teks ulasan yang telah dipraproses dan triplet opini hasil ekstraksi modul ekstraksi triplet opini. Kemudian, modul *postprocess* melakukan pengecekan pada tiap ekspresi aspek dan sentimen dari triplet opini untuk agar setiap ekspresi sesuai dengan kumpulan kosa kata teks masukan. Apabila terdapat kosa kata yang berbeda dari kumpulan kosa kata teks ulasan masukan, maka modul mencari kosa kata dari teks ulasan masukan dengan jarak terdekat dari kosa kata hasil prediksi.

Postprocess yang diterapkan pada tugas akhir memiliki dua variasi teknik untuk mengukur perbedaan antara dua teks. Variasi pertama menggunakan jarak levenshtein seperti yang diusulkan pada penelitian Zhang dkk., (2021). Untuk variasi kedua digunakan *cosine similarity* menggunakan vektor dari *embedding* model bahasa pralatih IndoT5 sebagai adaptasi dari penelitian Sitikhu dkk., (2019) yang menggunakan vektor *word embedding* pralatih untuk mengukur kesamaan antara dua teks. Untuk ekspresi dengan jumlah kata lebih dari 1, *postprocess* diterapkan pada tiap kata dalam ekspresi.

Terdapat perbedaan dalam penanganan *postprocess* pada ekstraksi triplet opini dengan penanganan aspek implisit. Apabila aspek yang diprediksi tidak terdapat pada kosa kata teks masukan, maka dicek terlebih dahulu apakah ekspresi yang dihasilkan adalah karakter “_” (NULL). Apabila ekspresi yang dihasilkan adalah NULL, maka *postprocess* diabaikan.



Gambar III.5 Ilustrasi Proses Perbaikan Prediksi dengan *Postprocess*

Ilustrasi proses perbaikan prediksi dengan *postprocess* dapat dilihat pada Gambar III.5. Ulasan “hotelnya bagus dan brsih” menghasilkan triplet opini (**hotel**, bagus, positif) dan (hotelnya, **bersih**, positif), dimana ekspresi aspek hotel dan ekspresi bersih tidak terdapat pada teks ulasan masukan. Untuk itu, dilakukan perbaikan pada kata paling mirip dengan mengukur jarak tiap kata dari teks ulasan masukan

dengan ekspresi tersebut. Untuk ekspresi hotel, kata paling mirip berdasarkan jarak levenshtein dan *cosine similarity* adalah “hotelnya”. Sementara itu untuk ekspresi bersih, kata paling mirip berdasarkan jarak levenshtein adalah “bersih” sedangkan kata paling mirip berdasarkan *cosine similarity* adalah “bagus”.

III.4 Desain Eksperimen

Eksperimen dilakukan pada tugas akhir untuk mendapatkan konfigurasi terbaik model ekstraksi triplet opini dengan pendekatan pembangkitan teks. Berdasarkan rancangan solusi yang diajukan, terdapat konfigurasi yang dapat dicari variasi terbaiknya, yaitu:

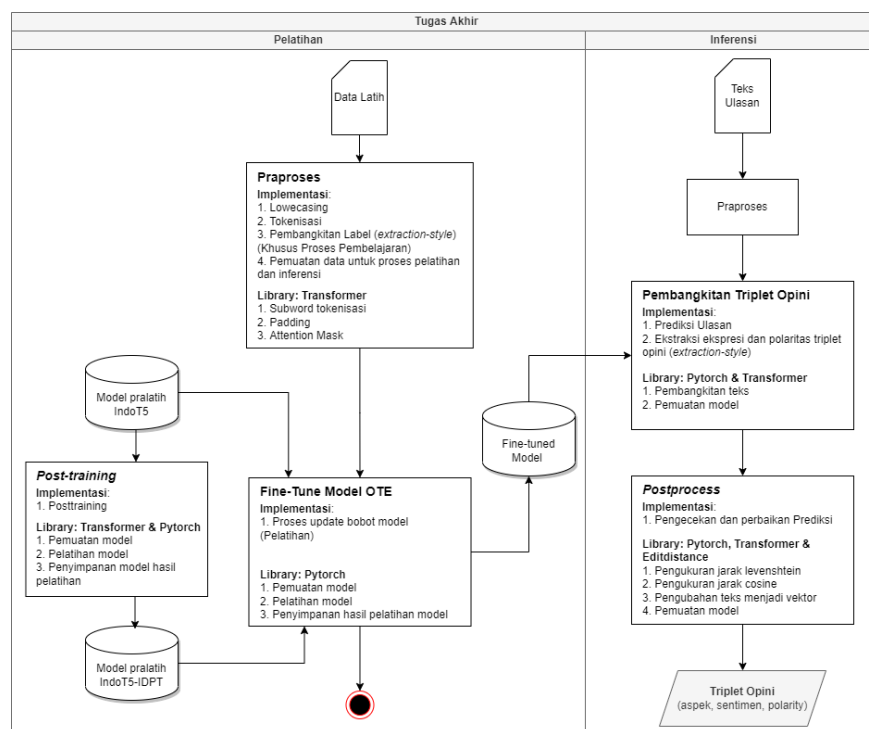
1. Melakukan *post-training* menggunakan data berdomain hotel pada IndoT5. Aditya (2022) menunjukkan peningkatan kinerja pada analisis sentimen berbasis aspek melalui *post-training* dengan data domain spesifik.
2. Metode perhitungan jarak antarkata pada *postprocess*. Zhang dkk., (2021) telah menunjukkan efektivitas penggunaan jarak levenshtein untuk mencari kata dengan jarak terdekat. Selain itu, Sithiku dkk., (2019) juga mengajukan penggunaan *cosine similarity* pada vektor representasi numerik teks untuk menghitung kesamaan antara data teks.
3. Jumlah kandidat (*number of beam*) yang digunakan saat proses pencarian untuk menghasilkan keluaran model bahasa. Raffel dkk., (2020) menemukan bahwa pada *task* dengan keluaran yang panjang, pencarian *beam* lebih baik ketimbang pencarian secara *greedy* (jumlah kandidat satu).

Eksperimen dilakukan dengan pencarian *grid search* untuk setiap kombinasi konfigurasi. Setiap konfigurasi dievaluasi menggunakan data validasi. Kemudian model tugas akhir dievaluasi menggunakan data uji dan dibandingkan kinerjanya terhadap model *baseline*, yaitu model yang diusulkan pada tugas akhir Wirawan (2021). Task yang dibandingkan terhadap model *baseline* adalah task ekstraksi triplet opini eksplisit dan ekstraksi triplet opini aspek implisit. Untuk metrik evaluasi yang digunakan adalah *F1-score* yang telah dijelaskan pada subbab II.6. Metrik ini digunakan agar kinerja model mempertimbangkan kasus ketika model

salah melakukan prediksi triplet opini dan kegagalan model untuk memprediksi triplet opini.

Teknik evaluasi yang digunakan penelitian tugas akhir ini adalah *exact match* yang digunakan pada penelitian tugas akhir Wirawan (2021) sebagai *baseline* penelitian. Artinya, keluaran dari model ekstraksi triplet opini harus tepat dan hasil prediksi yang bersifat benar sebagian akan dianggap gagal untuk diekstraksi atau salah. Sebagai contoh, “kamar mandi sangat bersih”, maka triplet opini yang dihasilkan haruslah “(kamar mandi, sangat bersih, positif)”. Hasil ekstraksi benar sebagian seperti “(kamar, sangat bersih, positif)” atau “(kamar mandi, bersih, positif)” tetap dianggap salah.

III.5 Implementasi



Gambar III.6 Detail Implementasi Solusi

Implementasi dari tugas akhir ini menggunakan bahasa pemrograman Python versi 3.7.12. Implementasi modul tugas akhir dilakukan dengan mengikuti rancangan solusi yang telah diuraikan pada subbab III.3. Detail implementasi dapat dilihat

pada Gambar III.6. Pelatihan dan pengujian model menggunakan platform Kaggle dengan menggunakan *hardware accelerator Graphics Processing Unit* atau GPU.

Berikut adalah penjelasan dari proses implementasi modul solusi tugas akhir.

1. Modul Praproses

Tokenisasi dalam modul praproses diimplementasikan dengan memanfaatkan *library* Transformer untuk mendapatkan tokenisasi subkata. Selain itu, proses penambahan *padding* dilakukan dengan memanfaatkan *library* Pytorch. Pada proses *fine-tuning* model, modul ini membangun target dengan format *extraction-style* agar keluaran model dapat diekstrak. Berikut adalah *pseudocode* untuk membangun target proses pelatihan dengan format *extraction-style*.

```
Func Generate Extraction Style Target(sentences, annotation):
    input: List of review text [sentences],
           List of index that refer triplet opinion expression [annotation]
    output: List of extraction-style triplet opinion
    extracted_targets <- []
    senttag2word <- {"POS": "positif", "NEG": "negatif", "NEU": "netral"}
    GENERAL_ASPECT <- '_'
    for i, label in enumerate(labels):
        all_triplet <- []
        for triplet in label:
            if triplet[0][0] == -1: # implicit aspect handling...
                aspect <- GENERAL_ASPECT
            else:
                start_idx, end_idx = tri[0][0], tri[0][-1]
                aspect = " ".join(sentences[i][start_idx : end_idx + 1])

                start_idx, end_idx = tri[1][0], tri[1][-1]
                sentiment <- " ".join(sentences[i][start_idx : end_idx + 1])
                polarity <- senttag2word[tri[2]]
                all_triplet.append((aspect, sentiment, polarity))
        # Build the extraction-style labels
        label_strs = ["(" + " ".join(l) + ")" for l in all_triplet]
        extracted_targets.append("; ".join(label_strs))
    return extracted_targets
END Func
```

2. Modul *Fine-Tune* Ekstraksi Triplet Opini

Implementasi modul *fine tuning* model dilakukan dengan memanfaatkan *library* Pytorch. Pembangunan model berbasiskan pada penggunaan model bahasa pralatih IndoT5 dengan memanfaatkan API Pytorch yang disediakan *library* Huggingface. Berikut adalah *pseudocode* untuk proses *fine-tuning* model ekstraksi triplet opini dengan format keluaran *extraction-style*.

```
Program Fine Tuning:
  input : model name, parameter model, dataset_path, path to save model
  output: fine-tuned model

  tokenizer, model <- init(parameter)  // initialize model configuration
  // initialize extraction-style target data
  train_data, val_data, test_data <- DataLoader(dataset_path)
  for epoch = 0 to n_epochs do  // training
    model.train()
    for batch in epoch do
      loss <- forward_propagation(batch)
      loss.backward() # Backward Propagation
      optimizer.step()
      optimizer.zero_grad()
    validation_per_epoch(val_data)
  model.save(path) //save model
END Program
```

3. Modul *Posttraining*

Implementasi modul *fine tuning* model dilakukan dengan memanfaatkan *library* Transformer dan Pytorch. Untuk melakukan *masking* pada data *post-training* sebagai objektif pelatihan tanpa supervisi digunakan modul *DataCollator* yang disediakan *library* Transformer untuk melanjutkan proses pelatihan tanpa supervisi dari model IndoT5. Berikut adalah *pseudocode* untuk proses *post-training* model IndoT5.

Program Posttraining:

```
input  : posttraining configuration (config)
        model parameter (params)
        in-domain dataset (data)

output : saved posttrained model

tokenizer, model <- init(params) // initialize model configuration
data_collator <- DataCollatorForT5MLM(tokenizer=tokenizer,config=config)
trainer <- Trainer(model=model,dataset=data,args=config,
                   tokenizer=tokenizer, data_collator=data_collator)
trainer.train(resume_from_checkpoint=checkpoint)
trainer.save_model() # Saves the tokenizer too for easy upload
validation(val_data)

END Program
```

4. Modul Pembangkitan Triplet Opini

Implementasi modul pembangkitan triplet opini dilakukan dengan memanfaatkan model bahasa pralatih yang telah dilatih (*fine-tune*) pada task ekstraksi triplet opini untuk membangkitkan triplet opini dari masukan kalimat ulasan yang memanfaatkan *library* Pytorch untuk memuat model dan untuk membangkitkan teks yang nantinya akan diekstrak sebagai triplet opini. Berikut ada *pseudocode* untuk proses ekstraksi elemen triplet opini dari hasil inferensi model.

Func Extract (sequence):

```
input : Text generated from model [sequence],
output: List of triplet opinions from model output
EMPTY, results <- '', []
triplets <- sequence.split("; ")
for elem in triplets:
    elem <- elem[1:-1] # Remove the in the start "(" and at the end ")".
    aspect, opinion, polarity <- elem.split(',')
    # Ignore problematic triplets
    if not problematic(aspect, opinion, polarity):
        results.append((a, b, c))
return results
```

END Func

5. Modul *Postprocess*

Modul *postprocess* digunakan untuk melakukan pengecekan dan perbaikan terhadap hasil pembangkitan triplet opini dengan kumpulan kosakata yang dibentuk dari kalimat ulasan masukan. Dalam langkah *postprocess*, modul bisa menerapkan proses normalisasi dengan berbagai strategi. Strategi normalisasi yang digunakan pada tugas akhir adalah tanpa normalisasi, normalisasi dengan jarak levenshtein dan normalisasi dengan jarak *cosine*. Pengukuran jarak levenshtein antar kata diimplementasikan dengan *library editdistance*, sementara *cosine similarity* diimplementasikan dengan menggunakan *library Pytorch*. Berikut adalah *pseudocode* dari langkah *postprocess*.

```
Func Postprocess(pair, sent, recover_term, implicit):
    input : Generated triplet opinion pair (extracted) [pair],
           Preprocessed sentence (list) [sent],
           Indicator to extract implicit aspect [implicit]
    output: Fixed triplet opinion pair
    sentiment_word_list <- ['positif', 'negatif', 'netral']
    new_pairs <- []
    for pair in pairs:
        at, ot, polarity = pair
    if at not in sents[i]: # Check aspect term
        if implicit and at in GENERAL_ASPECTS:
            at <- GENERAL_ASPECT
        else:
            at <- recover_term(at, sent)
    if ot not in sents[i]:
        ot <- recover_term(ot, sent)
    if polarity not in sentiment_word_list:
        polarity <- recover_term(polarity, sentiment_word_list)
        new_pairs.append((at, ot, polarity))
    return new_pairs
END Func
```

BAB IV

EKSPERIMEN DAN PENGUJIAN

IV.1 Eksperimen

IV.1.1 Tujuan Eksperimen

Tujuan dari eksperimen yang dilakukan dalam penelitian tugas akhir adalah sebagai berikut.

1. Melihat pengaruh penggunaan *post-training* terhadap kinerja model ekstraksi triplet opini eksplisit dan aspek implisit.
2. Melihat pengaruh perbedaan strategi normalisasi pada *postprocess* terhadap kinerja model ekstraksi triplet opini eksplisit dan aspek implisit.
3. Melihat pengaruh variasi jumlah *beam* (k) saat proses pencarian keluaran model bahasa pralatih dengan pendekatan pembangkitan teks terhadap kinerja ekstraksi triplet opini eksplisit dan aspek implisit.

IV.1.2 Data

Dataset yang digunakan untuk eksperimen dalam tugas akhir adalah kumpulan ulasan hotel yang diambil dari tugas akhir Wirawan (2021). Data terdiri dari 5000 teks ulasan dengan jumlah 76803 token, yang kemudian dibagi menjadi 3000 teks ulasan sebagai data latih, 1000 teks ulasan sebagai data validasi, dan 1000 teks ulasan sebagai data uji.

Pada tugas akhir ini terdapat dua jenis triplet opini yaitu triplet opini eksplisit dan triplet opini aspek implisit. Triplet opini eksplisit berarti triplet opini yang secara eksplisit dinyatakan dalam teks ulasan. Sedangkan triplet opini aspek implisit berarti triplet opini memiliki aspek yang dirujuk secara implisit oleh ekspresi sentimen (ekspresi sentimen tanpa pasangan) dan ditandai dengan karakter “_” sebagai aspek spesial NULL. Solusi tugas akhir untuk eksperimen pertama hanya menggunakan kalimat ulasan dengan triplet opini yang bersifat eksplisit saja,

sementara pada eksperimen kedua menggunakan kalimat ulasan dengan triplet opini eksplisit serta triplet opini dengan aspek implisit (campuran).

Tabel IV.1 Detail Dataset Triplet Opini Eksplisit

Data	Kalimat Total	Triplet Opini Eksplisit	Triplet Opini Aspek Implisit	Triplet Opini Total
Latih	3000	6251	0	6251
Validasi	1000	1989	0	1989
Uji	1000	2291	0	2291

Tabel IV.2 Detail Dataset Triplet Opini Campuran

Data	Kalimat Total	Triplet Opini Eksplisit	Triplet Opini Aspek Implisit	Triplet Opini Total
Latih	3000	6251	1290	7541
Validasi	1000	1989	467	2456
Uji	1000	2291	418	2709

Untuk mendukung tujuan eksperimen yang dilakukan pada tugas akhir, digunakan dua buah dataset yaitu dataset eksplisit dan dataset campuran. Perbedaan antara kedua data tersebut terletak dalam pengikutsertaan atau tidaknya triplet opini aspek implisit, dengan dataset eksplisit hanya mengikutsertakan triplet opini eksplisit sedangkan data campuran terdiri dari triplet opini eksplisit dan triplet opini aspek implisit. Eksperimen pertama menggunakan dataset eksplisit dan eksperimen kedua menggunakan dataset campuran. Detail dari kedua dataset eksperimen dapat dilihat pada Tabel IV.1 untuk ekstraksi triplet opini eksplisit dan Tabel IV.2 untuk ekstraksi triplet opini dengan penanganan aspek implisit.

Panjang rata-rata kalimat ulasan pada dataset adalah 15.72 kata, dan kalimat ulasan yang memiliki panjang lebih dari 40 hanya berjumlah 253 kalimat saja (~5% dari total). Berdasarkan data, dapat disimpulkan bahwa kalimat ulasan biasanya singkat karena kalimatnya langsung ke intinya, yaitu menyampaikan kesan terhadap

pengalaman menginap yang telah dialami dan tidak banyak pelanggan yang menulis panjang lebar untuk mengomentari setiap aspek hotel. Selain itu, data ulasan hotel tak berlabel berjumlah 136.788 kalimat digunakan untuk *post-training* model bahasa IndoT5.

IV.1.3 Skenario Eksperimen

Berdasarkan rancangan solusi yang telah dilakukan, beberapa konfigurasi yang dianggap dapat berpengaruh terhadap kinerja model dan diatur dalam eksperimen adalah sebagai berikut.

1. Penggunaan *post-training* menggunakan data berdomain hotel pada IndoT5. Proses *post-training* menggunakan data domain spesifik dapat berpengaruh terhadap representasi teks yang dihasilkan. IndoT5-IDPT adalah IndoT5 yang telah melalui proses *post-training* terhadap data spesifik domain hotel.
2. Metode perhitungan jarak antarkata pada *postprocess* yang diterapkan. Proses *postprocess* berpengaruh terhadap penanganan keluaran model yang tidak terdapat pada kalimat ulasan masukan.
3. Jumlah kandidat (*beam width*) yang digunakan saat proses pencarian untuk menghasilkan keluaran model bahasa.

Skenario eksperimen yang dilakukan dapat dilihat pada Tabel IV.3. Berdasarkan skenario tersebut, terdapat 12 varian model yang memiliki konfigurasi berbeda berdasarkan skenario yang dapat dilihat pada Tabel IV.4. Setiap varian model pada proses pencarian konfigurasi dievaluasi menggunakan data validasi untuk dipilih konfigurasi terbaiknya. Skenario eksperimen diterapkan pada 2 (dua) eksperimen dengan dataset yang telah dijelaskan pada subbab IV.1.2. Model dengan nilai evaluasi terbaik pada eksperimen pertama selanjutnya dievaluasi menggunakan data uji dan dibandingkan kinerjanya terhadap model *baseline*, yaitu model yang digunakan pada Tugas Akhir Wirawan (2021) untuk *task* ekstraksi triplet opini eksplisit.

Tabel IV.3 Skenario Eksperimen

ID	Tujuan Eksperimen	Parameter	Jumlah Variasi
P0	Melihat pengaruh penggunaan <i>post-training</i>	IndoT5	2
		IndoT5-IDPT	
P1	Melihat pengaruh perbedaan strategi normalisasi pada <i>postprocess</i>	Tanpa Normalisasi	3
		Jarak Levenshtein	
		<i>Cosine similarity</i>	
P2	Melihat pengaruh variasi <i>beam width</i> (k)	k=1	2
		k=5	
		Total	12

Penerapan *posttraining* pada skenario P0 dilakukan dengan melanjutkan proses pelatihan model IndoT5 pada task *masked language modelling* yang digunakan sebagai *task* pralatih model. Dampaknya adalah terdapat penyesuaian bobot pada tiap layer model. Dikarenakan tidak adanya kasus OOV untuk seluruh token kalimat ulasan pada data latih dan data validasi, maka tidak diterapkan penambahan kosa kata khusus untuk domain hotel sehingga tidak terdapat perubahan antara jumlah besar kumpulan kosa kata antara *tokenizer* pralatih dengan *tokenizer* model hasil *posttraining*.

Kemudian, penerapan metode perhitungan jarak kata dalam *postprocess* pada skenario P1 dilakukan untuk menyesuaikan keluaran model pendekatan pembangkitan teks agar memiliki perilaku yang sesuai dengan proses ekstraksi. Dampaknya adalah terdapat pengecekan dan perbaikan untuk kosa kata triplet opini yang tidak sesuai dengan kosa kata pada teks ulasan masukan sesuai dengan strategi yang dipilih. Variasi dari metode perhitungan jarak kata dalam *postprocess* yang dilakukan pada eksperimen ada 2, yaitu penggunaan jarak levenshtein atau *cosine similarity* pada vektor representasi numerik teks untuk mencari kosa kata paling mirip dengan keluaran model.

Terakhir, penerapan variasi jumlah kandidat yang dibangkitkan pada proses pencarian kalimat pada skenario P2 selanjutnya pada proses pembangkitan triplet

opini. Tujuan dari penambahan jumlah kandidat adalah mengurangi resiko dari kehilangan urutan token dengan probabilitas tinggi yang tersembunyi dengan tetap menyimpan jumlah hipotesis yang paling mungkin pada setiap *timestep* keluaran model sebanyak jumlah kandidat dan akhirnya memilih hipotesis yang memiliki probabilitas tertinggi secara keseluruhan. Terdapat 2 variasi jumlah kandidat yang diujikan, yaitu 1 dan 5.

Tabel IV.4 Konfigurasi Varian Model untuk Dataset Eksplisit dan Campuran

Model	Kinerja Ekstraksi Triplet Opini		
	<i>Pretrained LM</i> (P0)	<i>Postprocess</i> (P1)	Jumlah kandidat (P2)
M0	IndoT5	Tanpa Normalisasi	1
M1	IndoT5	Tanpa Normalisasi	5
M2	IndoT5	Jarak Levenshtein	1
M3	IndoT5	Jarak Levenshtein	5
M4	IndoT5	<i>Cosine Similarity</i>	1
M5	IndoT5	<i>Cosine Similarity</i>	5
M6	IndoT5-IDPT	Tanpa Normalisasi	1
M7	IndoT5-IDPT	Tanpa Normalisasi	5
M8	IndoT5-IDPT	Jarak Levenshtein	1
M9	IndoT5-IDPT	Jarak Levenshtein	5
M10	IndoT5-IDPT	<i>Cosine Similarity</i>	1
M11	IndoT5-IDPT	<i>Cosine Similarity</i>	5

Evaluasi dilakukan untuk ekstraksi triplet opini dengan teknik *exact match* seperti yang dijelaskan pada subbab III.4. Nilai hiperparameter yang digunakan dalam eksperimen dapat dilihat pada Tabel IV.5. Nilai *default* hiperparameter yang diatur merupakan nilai hiperparameter yang digunakan pada penelitian Zhang dkk., (2021).

Proses *posttraining* tersebut mencakup task *masked language* model yang merupakan pembelajaran tanpa supervisi (tidak memerlukan label). Konfigurasi yang digunakan untuk melakukan *posttraining* tersebut dapat dilihat pada Tabel

IV.6. *Posttraining* dilakukan dengan mengadaptasikan potongan kode yang tersedia dari huggingface untuk model bahasa pralatih T5.

Tabel IV.5 Nilai Hiperparameter Eksperimen

No	Hiperparameter	Nilai Default
1	<i>Batch size</i>	8
2	<i>Seed</i>	42
3	<i>Optimizer</i>	AdamW
4	Panjang kata maksimal	128 kata
5	Epochs	10
6	Number of Beam	1
7	Warmup steps	0
8	<i>Learning rate</i>	0.0003

Tabel IV.6 Konfigurasi *Post-training* IndoT5

No	Hiperparameter	Nilai Default
1	<i>Batch size</i>	8
2	<i>Seed</i>	42
3	epoch	10
4	Line by line	True
5	<i>Learning rate</i>	0.00005

IV.2 Hasil Eksperimen

IV.2.1 Hasil Eksperimen dengan Dataset Eksplisit

Tabel IV.7 Kinerja Validasi Skenario P0

Jenis Model Bahasa Pralatih	Kinerja Ekstraksi Triplet Opini Eksplisit		
	Precision	Recall	F1-score
IndoT5	0.8120	0.8480	0.8296
IndoT5-IDPT	0.7020	0.7896	0.7432

Hasil eksperimen untuk skenario P0 dapat dilihat pada Tabel IV.7. Berdasarkan hasil eksperimen tersebut, penerapan *post-training* pada IndoT5 (IndoT5-IDPT) menunjukkan penurunan kinerja (*F1-score*). Hal ini menunjukkan penerapan *post-training* untuk meningkatkan pemahaman terminologi domain perhotelan kurang tepat dilakukan pada solusi tugas akhir.

Tabel IV.8 Kinerja Validasi Skenario P1

Jenis <i>post-process</i>	Kinerja Ekstraksi Triplet Opini Eksplisit		
	Precision	Recall	F1-score
Tanpa Postprocess	0.8120	0.8480	0.8296
Jarak Levenshtein	0.8178	0.8541	0.8355
<i>Cosine Similarity</i>	0.8173	0.8535	0.8351

Hasil eksperimen untuk skenario P1 dapat dilihat pada Tabel IV.8. Penerapan tanpa *postprocess*, *postprocess* dengan jarak levenshtein, dan *postprocess* dengan *cosine similarity* memiliki kinerja *F1-score* secara berurut 0.8296, 0.8355 dan 0.8351. Hal ini menunjukkan adanya peningkatan kinerja dari penerapan *postprocess*, terutama pada penerapan normalisasi dengan memanfaatkan jarak levenshtein untuk mencari kosa kata paling mirip dengan keluaran model sebagai langkah *postprocess* dalam ekstraksi triplet opini eksplisit.

Berdasarkan hasil prediksi yang memerlukan *postprocess* pada data validasi, *postprocess* dengan jarak levenshtein memiliki keunggulan dibanding *cosine similarity* saat menangani ekspresi yang mengalami salah ketik (*typo*) seperti pada Tabel IV.9. Sementara itu, *postprocess* dengan *cosine* menggunakan vektor representasi numerik teks lebih unggul dalam menangani kasus ketika model membangkitkan teks berbeda dari label tetapi bermakna mirip. Contoh dari kasus tersebut dapat dilihat pada Tabel IV.10.

Tabel IV.9 Contoh Kasus *Typo* Hasil *Postprocess*

Teks Ulasan	Hasil Inferensi	
jangan menginap disini deh . lebih baik cari hotel lain . kelas deluxe tetapi seperti ekonomi sempit sekali , wc bau , colokkan listrik tidak ada . mau charger hp tidak bisa . nhabiskan uang saja . sangat kecewa .	Triplet Opini Sebenarnya	('colokkan listrik', 'tidak ada', 'negatif') ('wc', 'bau', 'negatif')
	Triplet Opini Hasil Prediksi	('colokkan listrik' (X), 'tidak ada', 'negatif') ('wc', 'bau', 'negatif')
	Jarak Levenshtein (“colokan”)	(jangan, 5), (menginap, 7), (disini, 7), (deh, 7), (lebih, 6), (baik, 6), (cari, 6), (hotel, 6), (lain, 5), (kelas, 5), (deluxe, 6), (tetapi, 7), (seperti, 7), (ekonomi, 6), (sempit, 7), (sekali, 6), (wc, 7), (bau, 6), (colokkan, 1), (listrik, 7), (tidak, 6), (ada, 6), (mau, 6), (charger, 6), (hp, 7), (tidak, 6), (bisa, 6), (nhabiskan, 7), (uang, 6), (saja, 6), (sangat, 6), (kecewa, 6).
	Hasil Postprocess (Levenshtein)	('colokkan listrik', 'tidak ada', 'negatif')
	Cosine Similarity (“colokan”)	(jangan, 0.43), (menginap, 0.37), (disini, 0.42), (deh, 0.45), (lebih, 0.48), (baik, 0.43), (cari, 0.46), (hotel, 0.45), (lain, 0.47), (kelas, 0.43), (deluxe, 0.42), (tetapi, 0.48), (seperti, 0.47), (ekonomi, 0.42), (sempit, 0.33), (sekali, 0.42), (wc, 0.39), (bau, 0.40), (colokkan, 0.51), (listrik, 0.47), (tidak, 0.50), (ada, 0.52), (mau, 0.45), (charger, 0.45), (hp, 0.40), (tidak, 0.50), (bisa, 0.48), (nhabiskan, 0.40), (uang, 0.44), (saja, 0.49), (sangat, 0.42), (kecewa, 0.34)
	Hasil Postprocess (<i>Cosine</i>)	('ada listrik' (X), 'tidak ada', 'negatif')

Tabel IV.10 Contoh Kasus Bukan *Typo* Hasil *Postprocess*

Teks Ulasan	Hasil Inferensi	
check in check out mudah . air hangat ada meski tidak begitu lancar . ada dispenser di setiap lantai . parkir luas . kamar tidak seluas seperti gambar . ada meja . tidak ada lemari . saluran tv tidak begitu jelas . lift dalam perbaikan . kamar mandi	Triplet Opini Sebenarnya	('air hangat', 'tidak begitu lancar', 'negatif')
	Triplet Opini Hasil Prediksi	('air dingin' (X), 'tidak begitu lancar', 'negatif')
	Jarak Levenshtein (“colokan”)	(check, 6), (in, 4), (check, 6), (out, 6), (mudah, 6), (air, 5), (hangat, 4), (ada, 6), (meski, 5), (tidak, 5), (begitu, 5), (lancar, 5), (ada, 6), (dispenser, 6), (di, 4), (setiap, 6), (lantai, 5), (parkir, 5), (luas, 6), (kamar, 6), (tidak, 5), (seluas, 6), (seperti, 7), (gambar, 6), (ada, 6), (

perlu ditingkatkan kebersihannya .		meja, 6), (tidak, 5), (ada, 6), (lemari, 6), (saluran, 6), (tv, 6), (tidak, 5), (begitu, 5), (jelas, 6), (lift, 5), (dalam, 5), (perbaikan, 7), (kamar, 6), (mandi, 4), (perlu, 6), (ditingkatkan, 7), (kebersihannya, 11)
	Hasil Postprocess (Levenshtein)	('air in' (X), 'tidak begitu lancar', 'negatif')
	Cosine Similarity ("colokan")	(check, 0.44), (in, 0.58), (check, 0.44), (out, 0.53), (mudah, 0.56), (air, 0.57), (hangat, 0.66), (ada, 0.59), (meski, 0.48), (tidak, 0.60), (begitu, 0.56), (lancar, 0.51), (ada, 0.59), (dispenser, 0.35), (di, 0.62), (setiap, 0.52), (lantai, 0.47), (parkir, 0.48), (luas, 0.53), (kamar, 0.50), (tidak, 0.60), (seluas, 0.43), (seperti, 0.56), (gambar, 0.51), (ada, 0.59), (meja, 0.51), (tidak, 0.60), (ada, 0.59), (lemari, 0.44), (saluran, 0.43), (tv, 0.43), (tidak, 0.60), (begitu, 0.56), (jelas, 0.54), (lift, 0.44), (dalam, 0.59), (perbaikan, 0.49), (kamar, 0.50), (mandi, 0.52), (perlu, 0.53), (ditingkatkan, 0.44), (kebersihannya, 0.47)
	Hasil Postprocess (Cosine)	('air hangat', 'tidak begitu lancar', 'negatif')

Terakhir, hasil eksperimen untuk skenario P2 dapat dilihat pada Tabel IV.11. Berdasarkan hasil eksperimen tersebut, penerapan variasi konfigurasi k (jumlah kandidat) tidak menunjukkan perubahan kinerja. Hal ini kemungkinan terjadi karena sebagian besar probabilitas untuk token berikutnya dari model bahasa dalam proses inferensi tidak memiliki urutan kata probabilitas tinggi yang tersembunyi, sehingga tidak perlu menyimpan jumlah hipotesis yang paling mungkin pada setiap *timestep* untuk memilih hipotesis yang memiliki probabilitas tertinggi secara keseluruhan.

Tabel IV.11 Kinerja Validasi Skenario P2

Jumlah kandidat	Kinerja Ekstraksi Triplet Opini Eksplisit		
	Precision	Recall	F1-score
1	0.8120	0.8480	0.8296
5	0.8120	0.8480	0.8296

Karena semakin besar nilai k yang digunakan semakin besar komputasi yang dilakukan. Maka, nilai k optimal yang dipilih adalah 1. Nilai parameter k yang dipilih sesuai dengan parameter *default* penelitian Zhang dkk., (2021).

Tabel IV.12 Detail Kinerja Variasi Model pada Dataset Eksplisit

Model	Kinerja Ekstraksi Triplet Opini Eksplisit		
	Precision	Recall	F1-score
M0	0.8120	0.8480	0.8296
M1	0.8120	0.8480	0.8296
M2	0.8178	0.8541	0.8355
M3	0.8178	0.8541	0.8355
M4	0.8173	0.8535	0.8351
M5	0.8173	0.8535	0.8351
M6	0.7020	0.7896	0.7432
M7	0.7020	0.7896	0.7432
M8	0.7110	0.7997	0.7527
M9	0.7110	0.7997	0.7527
M10	0.7083	0.7967	0.7499
M11	0.7083	0.7967	0.7499

IV.2.2 Hasil Eksperimen dengan Dataset Campuran

Hasil eksperimen untuk skenario P0 dapat dilihat pada Tabel IV.13. Berdasarkan hasil eksperimen tersebut, penerapan *post-training* pada IndoT5 (IndoT5-IDPT) menunjukkan penurunan kinerja (*F1-score*). Sama halnya dengan eksperimen ekstraksi triplet opini eksplisit, penerapan *post-training* untuk meningkatkan pemahaman terminologi domain perhotelan kurang tepat dilakukan pada pendekatan pembangkitan teks untuk ekstraksi triplet opini dengan aspek eksplisit.

Hasil eksperimen untuk skenario P1 dapat dilihat pada Tabel IV.14. Penerapan tanpa *postprocess*, *postprocess* dengan jarak levenshtein, dan *postprocess* dengan *cosine similarity* memiliki kinerja *F1-score* secara berurut 0.8326, 0.8354, dan 0.8350. Seperti eksperimen ekstraksi triplet opini eksplisit, penerapan *postprocess*

meningkatkan kinerja model terutama dengan normalisasi dengan jarak levenshtein untuk mencari kosa kata paling mirip dengan keluaran model.

Tabel IV.13 Kinerja Validasi Skenario P0

Jenis Model Bahasa Pralatih	Kinerja Ekstraksi Triplet Opini Aspek Implisit		
	Precision	Recall	F1-score
IndoT5	0.8064	0.8606	0.8326
IndoT5-IDPT	0.7925	0.8529	0.8216

Tabel IV.14 Kinerja Validasi Skenario P1

Jenis <i>postprocess</i>	Kinerja Ekstraksi Triplet Opini Aspek Implisit		
	Precision	Recall	F1-score
Tanpa Postprocess	0.8064	0.8606	0.8326
Jarak Levenshtein	0.8091	0.8635	0.8354
<i>Cosine Similarity</i>	0.8087	0.8631	0.8350

Tabel IV.15 Kinerja Variasi Skenario P2

Jumlah kandidat	Kinerja Ekstraksi Triplet Opini Aspek Implisit		
	Precision	Recall	F1-score
1	0.8064	0.8606	0.8326
5	0.8064	0.8606	0.8326

Terakhir, hasil eksperimen untuk skenario P2 dapat dilihat pada Tabel IV.15. Berdasarkan hasil eksperimen tersebut, penerapan variasi konfigurasi k (jumlah kandidat) tidak menunjukkan perubahan kinerja (*F1-score*). Sama halnya dengan eksperimen ekstraksi triplet opini eksplisit, hal ini kemungkinan terjadi karena sebagian besar probabilitas untuk token berikutnya dari model bahasa dalam proses inferensi tidak memiliki urutan kata probabilitas tinggi yang tersembunyi, sehingga tidak perlu menyimpan jumlah hipotesis yang paling mungkin pada setiap timestep untuk memilih hipotesis yang memiliki probabilitas tertinggi secara keseluruhan.

Karena semakin besar nilai k yang digunakan semakin besar langkah komputasi yang dilakukan, maka, nilai k optimal yang dipilih adalah 1.

Tabel IV.16 Detail Kinerja Variasi Model pada Dataset Campuran

Model	Kinerja Ekstraksi Triplet Opini Aspek Implisit		
	Precision	Recall	F1-score
M0	0.8064	0.8606	0.8326
M1	0.8064	0.8606	0.8326
M2	0.8091	0.8635	0.8354
M3	0.8091	0.8635	0.8354
M4	0.8087	0.8631	0.8350
M5	0.8087	0.8631	0.8350
M6	0.7925	0.8529	0.8216
M7	0.7925	0.8529	0.8216
M8	0.7952	0.8557	0.8243
M9	0.7952	0.8557	0.8243
M10	0.7944	0.8549	0.8235
M11	0.7944	0.8549	0.8235

IV.3 Analisis Hasil Pengujian dengan Dataset Eksplisit

Pengujian dilakukan terhadap data uji menggunakan model yang dilatih dengan konfigurasi terbaik yang diperoleh dari eksperimen yang telah dilakukan terhadap data validasi untuk mengukur dan membandingkan kinerja solusi tugas akhir dengan model penelitian sebelumnya (baseline) pada ekstraksi triplet opini eksplisit.. Model yang dibangun untuk task ekstraksi triplet opini eksplisit dibandingkan dengan model *baseline*, yaitu model GTS yang digunakan pada tugas akhir Wirawan (2021).

Konfigurasi parameter yang digunakan untuk model *baseline* mengikuti konfigurasi terbaik pada penelitian Wirawan (2021). Berdasarkan hasil eksperimen yang dijelaskan pada subbab IV.2.1, didapatkan konfigurasi terbaik model Tugas Akhir

(TA) ekstraksi triplet opini eksplisit yaitu menggunakan IndoT5, *postprocess* dengan jarak levenshtein dan k (jumlah kandidat) yang digunakan adalah 1.

Pengujian untuk ekstraksi triplet opini eksplisit mencakup perbandingan dengan model tugas akhir Wirawan (2021). Model tugas akhir Wirawan (2021) menggunakan model GTS dengan memanfaatkan IndoBERT dan konfigurasi yang digunakan adalah konfigurasi terbaik pada penelitian tersebut, yaitu dengan parameter data latih menggunakan data latih *uncleaned*, model bahasa yang digunakan adalah model bahasa monolingual, dan pemanfaatan model bahasanya secara *fine-tuning*.

Tabel IV.17 Hasil Pengujian Model Terbaik Dibandingkan Model Baseline untuk Ekstraksi Triplet Opini Eksplisit

Model	Kinerja Ekstraksi Triplet Opini Eksplisit		
	Precision	Recall	F1-score
GTS-GW	0.8573	0.7039	0.7731
TA	0.8516	0.7753	0.8117

Hasil pengujian model terbaik dan perbandingannya dengan model *baseline* dapat dilihat pada Tabel IV.17. Model tugas akhir (TA) adalah model dengan konfigurasi terbaik yaitu varian M2, sedangkan model GTS-GW adalah model pada penelitian tugas akhir Wirawan (2021). Berdasarkan hasil pengujian tersebut, kinerja model tugas akhir memiliki kinerja yang lebih tinggi dibanding model GTS-GW dengan perbedaan sekitar 0.04.

Hasil model dengan pendekatan pembangkitan teks memiliki nilai *recall* yang lebih tinggi dibanding model GTS-GW. Artinya, model TA memiliki kegagalan yang lebih sedikit dalam mengekstraksi triplet opini dibanding model GTS-GW. Hal ini dapat menjadi kelebihan dari penggunaan pendekatan pembangkitan teks yaitu mampu mendeteksi triplet opini lebih baik ketimbang model dengan pendekatan *sequence labelling*.

Untuk menentukan apakah model TA lebih baik ketimbang model GTS-GW dilakukan uji statistik McNemar dengan nilai signifikan 0.05 (*alpha*). Uji yang dilakukan menggunakan *one-tailed test* untuk mengetahui apakah kinerja model TA lebih baik

ketimbang model GTS-GW dengan probabilitas 0.05. Maka dari itu, hipotesis *null* (H_0) dari uji statistik adalah kedua model memiliki kinerja yang sama (*error rate* yang sama). Data yang digunakan untuk uji hipotesis dapat dilihat pada Tabel IV.18.

Tabel IV.18 Ringkasan Hasil Pengujian Model

	TA		
GTS-GW	Benar	Gagal	Total
Benar	1482	95	1577
Gagal	287	427	714
Total	1769	522	2291

Dari hasil nilai uji statistik McNemar, diperoleh nilai statistik McNemar (χ^2) sebesar 95.5 dengan nilai *p-value* sebesar 1.49×10^{-22} . Artinya, hipotesis *null* (H_0) yang menyatakan bahwa kedua model memiliki kinerja yang sama dapat ditolak karena $\chi^2 > \chi^{0.05}_{(5.43)}$ pada *one tailed test* untuk uji statistic McNemar berdasarkan nilai signifikan (*p-value*) yang telah ditetapkan.

Berdasarkan analisis terhadap prediksi model TA, model dapat menangani kasus ulasan yang mengandung bahasa lain (*code-mixing*) dengan baik. Selain itu, model gagal atau salah mengekstraksi 628 pasangan triplet opini pada pengujian yang dilakukan terhadap data uji. Analisis terhadap prediksi dari model TA sebagai berikut.

1. Model TA sering mengalami kesalahan prediksi pada kalimat ulasan yang letak ekspresi sentimen yang berdekatan dengan ekspresi aspek bukan pasangannya, terutama pada ulasan dengan ekspresi sentimen tanpa pasangan. Hal ini kemungkinan terjadi karena model salah mengartikan maksud dari ulasan yang ambigu sehingga salah mengekstrak relasi dari ekspresi aspek dan sentimen yang berdekatan. Contoh dari kasus tersebut dapat dilihat pada Tabel IV.19.

Tabel IV.19 Contoh Kasus Kesalahan Umum 1

Kalimat	Triplet Sebenarnya	Prediksi TA
tempatnya strategis , pelayanannya baik , nyaman dan lengkap .	('tempatnya', 'strategis', 'positif') ('pelayanannya', 'baik', 'positif')	('tempatnya', 'strategis', 'positif') ('pelayanannya', 'baik', 'positif') ('pelayanannya', 'nyaman', 'positif') (X) ('pelayanannya', 'lengkap', 'positif') (X)
tempatnya bersih dan nyaman , good , like it .	('tempatnya', 'bersih', 'positif') ('tempatnya', 'nyaman', 'positif')	('tempatnya', 'bersih', 'positif') ('tempatnya', 'nyaman', 'positif') ('tempatnya', 'good', 'positif') (X)

2. Model TA sering mengalami kegagalan untuk memprediksi triplet opini dengan ekspresi aspek berupa kata sifat. Hal ini kemungkinan terjadi karena kebanyakan triplet opini yang ada pada data latih merupakan kata benda (entitas) seperti “kamar”, “hotel”, dan lainnya, sehingga model mengalami kesulitan untuk mengekstraksi triplet opini dengan ekspresi aspek berupa kata kerja seperti “menginap”. Contoh dari kasus tersebut dapat dilihat pada Tabel IV.20.

Tabel IV.20 Contoh Kasus Kesalahan Umum 2

Kalimat	Triplet Sebenarnya	Prediksi TA
menyenangkan menginap disini .	('menginap', 'menyenangkan', 'positif')	(X)
saya senang menginap disini , nyaman .	('menginap', 'nyaman', 'positif')	(X)

Selain itu, dilakukan studi kasus pada 129 triplet opini hasil prediksi yang hanya memiliki perbedaan ekspresi aspek atau ekspresi sentimen dengan triplet opini pada

data uji. Dari studi kasus tersebut, ditemukan 37 triplet opini dengan ekspresi aspek yang sedikit berbeda dengan label dan 36 triplet opini dengan ekspresi sentimen yang sedikit berbeda dengan label namun masih memiliki makna yang sama. Contoh dari kasus tersebut dapat dilihat pada Tabel IV.21.

Tabel IV.21 Contoh Kasus Kesalahan Akibat Sedikit Perbedaan pada Ekspresi dalam Triplet Opini

Kalimat	Triplet Sebenarnya	Prediksi TA
ranjang tidur yang sudah mulai rusak .	('ranjang tidur', 'rusak', 'negatif')	('ranjang tidur', 'mulai rusak' (X), 'negatif')
keadaan kamar tidak sesuai dengan foto yang ada .	('keadaan kamar', 'tidak sesuai', 'negatif')	('kamar' (X), 'tidak sesuai', 'negatif')
free wifi mengecewakan karena tidak bisa digunakan instalasi listrik kurang rapi .	('free wifi', 'mengecewakan', 'negatif')	('wifi' (X), 'mengecewakan', 'negatif')

Studi kasus lainnya dilakukan pada kasus kalimat pernyataan ataupun saran yang tidak mengandung ekspresi sentimen. Terdapat 32 teks ulasan dengan kasus tersebut dan model salah memprediksi sebanyak 7 pasangan triplet opini pada 6 teks ulasan. Beberapa contoh kesalahan pada kasus tersebut dapat dilihat pada Tabel IV.22.

Tabel IV.22 Contoh Kasus Kesalahan pada Kalimat Saran atau Pernyataan

Kalimat	Triplet Sebenarnya	Prediksi TA
kalau bisa disediakan sandal untuk tamu .	-	('sandal', 'disediakan', 'positif') (X)
pihak hotel tidak memberikan keterangan yang lengkap tentang persyaratan menginap pada airy .	-	('keterangan', 'tidak lengkap', 'negatif') (X)

IV.4 Analisis Hasil Pengujian dengan Dataset Campuran

Pengujian dilakukan terhadap data uji menggunakan model yang dilatih dengan konfigurasi terbaik yang diperoleh dari eksperimen yang telah dilakukan terhadap data validasi untuk mengukur kinerja solusi tugas akhir pada ekstraksi triplet opini dengan penanganan aspek implisit. Berdasarkan hasil eksperimen yang dijelaskan pada subbab IV.2.2, konfigurasi terbaik model tugas akhir ekstraksi triplet opini aspek implisit yaitu menggunakan IndoT5, *postprocess* dengan jarak levenshtein dan k (jumlah kandidat) yang digunakan adalah 1. Metrik yang digunakan pada pengujian ini adalah nilai *F1-score* model terhadap data uji. Hasil kinerja model terhadap data uji dapat dilihat pada Tabel IV.23.

Tabel IV.23 Hasil Pengujian Model Terbaik untuk Ekstraksi Triplet Opini dengan Aspek Implisit

Model	Kinerja Ekstraksi Triplet Opini dengan Aspek Implisit		
	Precision	Recall	F1-score
TA	0.8200	0.7715	0.7950

Model gagal atau salah mengekstraksi 748 pasangan triplet opini ketika diujikan dengan data uji. Untuk pasangan triplet opini eksplisit, analisis pada prediksi model menunjukkan perilaku seperti yang telah dijelaskan pada subbab IV.3. Analisis terhadap prediksi model untuk pasangan triplet opini dengan aspek implisit sebagai berikut.

1. Terdapat pasangan triplet opini dengan aspek implisit yang ditandai dengan aspek spesial (NULL) yang diekstrak sebagai pasangan triplet opini eksplisit dengan ekspresi aspek yang berdekatan dengan ekspresi sentimen tanpa pasangan. Sama seperti kasus pada ekstraksi triplet opini eksplisit, hal ini kemungkinan terjadi karena model salah mengartikan maksud dari ulasan yang ambigu sehingga salah mengekstrak relasi dari ekspresi aspek dan sentimen yang berdekatan. Contoh dari kasus tersebut dapat dilihat pada Tabel IV.24.

Tabel IV.24 Contoh Kasus Kesalahan Umum 1 Triplet Opini Aspek Implisit

Kalimat	Triplet Sebenarnya	Prediksi TA
pokok nya puas , tenang suasana nya .	('_', 'puas', 'positif') ('suasana nya', 'tenang', 'positif')	('suasana nya' (X), 'puas', 'positif') ('suasana nya', 'tenang', 'positif')
pelayanan ramah , bersih dan fasilitas sesuai , cuma wifi agak lambat saja . lokasi dekat kemanakemana .	('pelayanan', 'ramah', 'positif') ('_', 'bersih', 'positif') ('fasilitas', 'sesuai', 'positif') ('wifi', 'agak lambat', 'negatif')	('pelayanan', 'ramah', 'positif') ('pelayanan' (X), 'bersih', 'positif') ('fasilitas', 'sesuai', 'positif') ('wifi', 'agak lambat', 'negatif')

2. Terdapat pasangan triplet opini eksplisit yang diprediksi menjadi pasangan triplet opini dengan aspek implisit. Hal ini kemungkinan terjadi karena ekstraksi triplet opini yang dilakukan menggabungkan triplet opini eksplisit dan triplet opini dengan aspek implisit sehingga aspek dalam triplet opini eksplisit yang jarang ditemukan pada data latih diprediksi sebagai aspek spesial (NULL). Contoh dari kasus tersebut dapat dilihat pada Tabel IV.25.

Tabel IV.25 Contoh Kasus Kesalahan Umum 2 Triplet Opini Aspek Implisit

Kalimat	Triplet Sebenarnya	Prediksi TA
tidak terlalu jauh dari pusat kota .	('dari pusat kota', 'tidak terlalu jauh', 'positif')	('_' (X), 'tidak terlalu jauh', 'positif')
nyaman stay disini mas yang jaga malam baik .	('mas yang jaga malam', 'baik', 'positif') ('stay disini', 'nyaman', 'positif')	(yang jaga malam' (X), 'baik', 'positif') ('_' (X), 'nyaman', 'positif')

Berdasarkan hasil analisis dari kasus-kasus umum tersebut, kasus kesalahan dan kegagalan umum yang terjadi pada ekstraksi triplet opini dengan aspek implisit terjadi sebagai akibat dari penambahan kompleksitas dari *task* yang dilakukan, yaitu ekstraksi triplet opini dengan penanganan pada aspek implisit. Meskipun penambahan kompleksitas *task* yang dilakukan menambahkan kasus kesalahan yang dapat terjadi, jumlah kesalahan pada kasus umum yang terjadi pada ekstraksi triplet opini eksplisit menjadi berkurang. Hal ini dapat terjadi karena penambahan data ekspresi sentimen pada triplet opini aspek implisit yang dapat dipelajari model pada data latih membuat model lebih baik dalam mengekstraksi triplet opini pada data uji.

IV.5 Analisis Pengaruh *Posttraining* pada Ekstraksi Triplet Opini

Penerapan *posttraining* dilakukan untuk meningkatkan pemahaman model bahasa pralatih terhadap terminologi domain perhotelan. Hal ini berpengaruh terhadap bobot model bahasa pralatih. Perubahan bobot mengubah representasi teks yang dihasilkan. Sebagai contoh, pada kalimat ulasan “tempat lumayan bagus kamar bersih dan wangi dekat dengan tempat jual makanan recommend” memiliki hasil keluaran layer *encoder* yang berbeda antara model bahasa IndoT5 dengan IndoT5-IDPT yang dapat dilihat pada Tabel IV.26.

Tabel IV.26 Perbedaan Representasi Sebelum dan Sesudah *Post-training*

Kalimat	tempat lumayan bagus kamar bersih dan wangidekat dengan tempat jual makanan recommend
Representasi Encoder IndoT5 (Pooled)	[-0.1284, -0.3514, -0.3277, 0.1183, -0.1369, -0.1547, -0.0334, -0.2040, 0.3408, 0.0623, -0.1284, 0.0369, 0.0053, 0.1111, 0.0713, 0.7388, -0.1067, 0.1120]
Representasi Encoder IndoT5-IDPT (Pooled)	[-0.1287, -0.3517, -0.3279, 0.1185, -0.1369, -0.1546, -0.0332, -0.2043, 0.3406, 0.0622, -0.1287, 0.0370, 0.0051, 0.1113, 0.0712, 0.7387, -0.1067, 0.1133]

Kinerja dari model bahasa IndoT5 lebih tinggi ketimbang IndoT5-IDPT untuk ekstraksi triplet opini eksplisit. Analisa perbedaan hasil prediksi kedua varian model dengan konfigurasi *default* untuk parameter lainnya (M0 vs M6)

menunjukkan model IndoT5-IDPT lebih peka terhadap ekspresi sentimen dibanding IndoT5 yang menyebabkan model sering salah mengekstraksi pasangan triplet opini dibanding IndoT5. Contoh dari kasus kesalahan yang terjadi pada model IndoT5-IDPT dapat dilihat pada Tabel IV.27.

Tabel IV.27 Contoh Kasus Kesalahan Akibat *Posttraining*

Teks Ulasan	Hasil Inferensi	
nyaman dan bersih.	Triplet Sebenarnya	-
	Prediksi M0 (TA)	-
	Prediksi M6 (TA)	('kamar', 'nyaman', 'positif') (X) ('kamar', 'bersih', 'positif') (X)

Untuk ekstraksi triplet opini dengan aspek implisit perbedaan kinerja IndoT5 dengan IndoT5-IDPT tidak jauh berbeda. Seperti yang telah dijelaskan sebelumnya, hal ini kemungkinan terjadi karena model IndoT5-IDPT lebih peka terhadap ekspresi sentimen dibanding IndoT5 sehingga saat triplet opini dengan aspek implisit dari ekspresi sentimen tanpa pasangan disertakan dalam data uji, kedua model memiliki kinerja yang tidak banyak berbeda.

BAB V

KESIMPULAN DAN SARAN

V.1 Kesimpulan

Kesimpulan dari penelitian Tugas Akhir ini adalah sebagai berikut.

1. Untuk membangun model ekstraksi triplet opini dengan pendekatan pembangkitan teks diperlukan korpus bahasa Indonesia dengan format *extraction-style* untuk melatih model bahasa pralatih IndoT5 serta *postprocess* untuk memperbaiki kesalahan pada hasil prediksi yang tidak terdapat pada teks masukan.
2. Penanganan terhadap pasangan triplet opini dengan aspek implisit dapat dilakukan dengan memasang ekspresi sentimen tanpa pasangan dengan aspek spesial (NULL). *F1-score* dari hasil pengujian yang didapatkan dari model dengan konfigurasi terbaik tugas akhir untuk ekstraksi triplet opini dengan aspek implisit adalah 0.7950.
3. Kinerja model tugas akhir lebih baik secara signifikan dibandingkan model *baseline* (GTS-GW) berdasarkan uji statistik McNemer. Kinerja *F1-score* dari hasil pengujian yang didapatkan model terbaik tugas akhir untuk ekstraksi triplet opini eksplisit adalah 0.8117, sementara kinerja model *baseline* (GTS-GW) dengan konfigurasi terbaik adalah 0.7731.
4. Berdasarkan analisis, limitasi dari model ekstraksi triplet opini dengan pendekatan pembangkitan teks adalah sering mengalami kesalahan prediksi apabila terdapat ekspresi sentimen tanpa pasangan aspek dan sering mengalami kegagalan prediksi pada triplet opini dengan ekspresi aspek berupa kata sifat. Penambahan penanganan aspek implisit juga membuat model sering memprediksi aspek yang jarang ditemukan pada data latih diprediksi sebagai aspek spesial (NULL). Evaluasi kesalahan juga menunjukkan model dapat membangkitkan triplet opini dengan ekspresi

aspek atau sentimen yang sedikit berbeda dari ekspresi dalam teks masukan meskipun masih memiliki makna yang sama.

V.2 Saran

Saran-saran terkait tugas akhir yang dapat diteliti lebih lanjut adalah sebagai berikut.

1. Menggunakan variasi model bahasa pralatih lain dengan jumlah parameter yang lebih besar seperti *IndoT5-large* untuk meningkatkan kinerja model ekstraksi triplet opini. Penelitian Raffel dkk., (2020) membuktikan peningkatan jumlah parameter model dapat meningkatkan kinerja model. Meskipun demikian, hal ini perlu dibarengi dengan peningkatan *resource* komputasi untuk melatih model yang tidak dapat diakomodasi pada penelitian ini.
2. Menggunakan variasi *word embedding* lain seperti *word2vec* yang dilatih dengan korpus bahasa Indonesia dalam jumlah besar (seperti korpus Wikipedia) dengan penambahan formalisasi kata informal dalam teks ulasan untuk langkah *postprocess*.
3. Variasi pencarian kata paling mirip dalam *postprocess* untuk *task* ekstraksi dengan pendekatan pembangkitan teks dapat ditelusuri lebih lanjut. Salah satu pengembangan yang dapat dilakukan adalah dengan memperhatikan konteks kalimat saat mengukur kemiripan antara 2 kata dalam *postprocess*. Penelitian ini hanya membahas pengukuran kemiripan kata secara leksikal dan berdasarkan representasi vektor dari kata tanpa memerhatikan konteks kalimat.

DAFTAR REFERENSI

- Aditya, I. P. E. S. (2022). Pembelajaran Transfer Dengan Post Training Untuk Analisis Sentimen Berbasis Aspek Berbahasa Indonesia. Laporan Tesis, Institut Teknologi Bandung, Bandung.
- Chen Zhang, Qiuchi Li, Dawei Song, and Benyou Wang. 2020. A Multi-task Learning Framework for Opinion Triplet Extraction. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 819–828, Online. Association for Computational Linguistics.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, & Peter J. Liu (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21(140), 1-67.
- Genadi, R. A. (2021). Pembelajaran Transfer dan Representasi Berbasis Span Untuk Ekstraksi Triplet Opini Untuk Analisis Sentimen Berbasis Aspek. Laporan Tesis, Institut Teknologi Bandung, Bandung.
- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-Domain Targeted Sentiment Analysis via Span-Based Extraction and Classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546, Florence, Italy. Association for Computational Linguistics.
- Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.
- Peng, Haiyun & Xu, Lu & Bing, Lidong & Huang, Fei & Lu, Wei & Si, Luo. (2020). Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*. 34. 8600-8607. 10.1609/aaai.v34i05.6383.
- Sitikhu, Pinky & Pahi, Kritish & Thapa, Pujan & Shakya, Subarna. (2019). A Comparison of Semantic Similarity Methods for Maximum Human Interpretability.
- Qi Su, Xinying Xu, Honglei Guo, Zhili Guo, Xian Wu, Xiaoxun Zhang, Bin Swen, and Zhong Su. 2008. Hidden sentiment association in chinese web opinion mining. In *Proceedings of the 17th international conference on World Wide Web (WWW '08)*. Association for Computing Machinery, New York, NY, USA, 959–968. <https://doi.org/10.1145/1367497.1367627>
- Rendyanto, H. (2021). Model Bahasa Generatif Untuk Analisis Sentimen Berbasis Aspek. Laporan Tesis, Institut Teknologi Bandung, Bandung.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing and Wai Lam (2021). Towards Generative Aspect-Based Sentiment Analysis.
- Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing and Wai Lam (2022). A Survey on Aspect Based Sentiment Analysis: Tasks, Methods, and Challenges. 10.48550/ARXIV.2203.01054.
- Wirawan, G. P. (2021). Ekstraksi Triplet Opini Pada Analisis Sentimen Berbasis Aspek dengan Pendekatan Grid Tagging Scheme. Laporan Tugas Akhir, Institut Teknologi Bandung, Bandung.
- Xue, C. (2021). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 483–498). Association for Computational Linguistics.
- Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 2576–2585, Online. Association for Computational Linguistics.
- Zhao, H., Huang, L., Zhang, R., & Lu, Q. (2020, July). SpanMlt: A Span-based Multi-Task Learning Framework for Pair-wise Aspect and Opinion Terms Extraction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 3239-3248).

Lampiran A. Perbaikan Anotasi Triplet Opini

A.1 Perbaikan Anotasi Polaritas Triplet Opini

Tabel A.1 Daftar Perbaikan Anotasi Polaritas Triplet Opini

No.	Ulasan	Triplet Opini	Triplet Opini seharusnya
1.	kamar mandi bau banget , kamar gelap , overall oke untuk harga segitu . tolong di perhatikan lampu dan kamar mandi nya iya . thankyou !	(kamar mandi, bau banget, positif)	(kamar mandi, bau banget, negatif)
2.	breakfast biasa saja dan porsinya sedikit .	(breakfast, biasa saja, negatif)	(breakfast, biasa saja, netral)
3.	kamar bersih , kamar mandi bersih , kolam renang baik , classic atmosphere .	(kamar, bersih, negatif)	(kamar, bersih, positif)
4.	pelayan cepat .	(pelayan, cepat, negatif)	(pelayan, cepat, positif)
5.	nah ini tempat walaupun minimalis tetapi nyaman . ac nya dingin , staf ramah . sudah langganan juga he .	(tempat, minimalis, positif)	(tempat, minimalis, netral)
6.	diluar perkiraan , harganya murah tetapi mendapatkan kamar yang besar dan fasilitas yang lengkap . sukses terus untuk traveloka dan airy , pahala kalian banyak karena selalu memudahkan urusan banyak orang .	(harganya, murah, negatif)	(harganya, murah, positif)
7.	hotelnya biasa biasa saja , sesuai budget . tetapi kemarin kena cas 100 ribu gara gara yang menginap 3 or .	(hotelnya, biasa biasa saja, negatif)	(hotelnya, biasa biasa saja, netral)
8.	tetapi air sering mati itu saja yang lain boleh sama remote ac dan tv tidak ada .	(remote ac, tidak ada, positif)	(remote ac, tidak ada, negatif)
9.	interior room keren , room bersih , hallway kotor bekas makanan tidak diangkat . kalau yang penakut agak seram pemandangan kuburan . breakfast biasa saja . dekat ke kota tetapi super macet di jam sibuk .	(breakfast, biasa saja, negatif)	(breakfast, biasa saja, netral)
10.	kurang bersih saja . tetapi keseluruhan lumayan lah .	(keseluruhan, lumayan, negatif)	(keseluruhan, lumayan, positif)
11.	kondisi kamar standar banget . saya sering menggunakan airy rooms . tetapi kali ini yang paling tidak nyaman . 1 . ac kurang dingin 2 . tidak ada snack 3 . heater water tidak berfungsi 4 . kamar biasa saja not recommended sih , kalau menurut saya .	(kamar, biasa saja, negatif)	(kamar, biasa saja, netral)

12.	tempatya sederhana tetapi cukup bersih , klau malam banyak suara jaangkrik dan kambing , suasana desa , mobil langsung parkir depan kamar , air lancar , air panas ada , kamar mandi agaak bau (dari saluran air) .	(tempatnya, sederhana, positif)	(tempatnya, sederhana, netral)
13.	gedungnya perlu direnovasi , kamar terlihat kusam begitu juga depan furniturnya dan banyak nyamuk . menu breakfast sederhana , tetapi rasanya enak .	(breakfast, sederhana, negatif)	(breakfast, sederhana, netral)

A.2 Perbaikan Anotasi Pasangan Triplet Opini

Terdapat 81 pasangan triplet opini yang diperbaiki. Triplet opini diperoleh dari ekspresi sentimen yang sebelumnya tidak memiliki pasangan ekspresi aspek eksplisit pada dataset, namun seharusnya memiliki pasangan ekspresi aspek. Contoh dari perbaikan anotasi pasangan triplet opini eksplisit dapat dilihat pada Tabel A.2. Detail perbaikan anotasi pasangan triplet opini eksplisit secara lebih lengkap dapat dilihat pada tautan berikut (https://docs.google.com/spreadsheets/d/15p5DYbUxRl33lmkO2fxZsSCKz_9GPus8uHcGzqmw7tk/edit?usp=sharing).

Tabel A.2 Contoh Perbaikan Label Triplet Opini

No.	Ulasan	Triplet Opini	Triplet Opini setelah diperbaiki
1.	kamarnya bersih , tidak mengecewakan , strategis untuk kemanamana .	(kamarnya, bersih, positif)	(kamarnya, bersih, positif) (kamarnya, tidak mengecewakan, netral)
2.	baik , pelayanan , nya ramah dan siap siaga .	(pelayanan, ramah, positif) (pelayanan, siap siaga, positif)	(pelayanan, ramah, positif) (pelayanan, siap siaga, positif) (pelayanan, baik, positif)
3.	kali ini dapat kamar dekat kolam renang , anak-anak senang , menu sarapan sudah sesuai keinginan , sipp .	-	(menu sarapan, sudah sesuai keinginan, positif)
4.	kamar budgett sekali di jakarta .	-	(kamar, budgett sekali, negatif)
5	kamarnya dicat ulang , kalau perlu didekor ulang biar kelihatan hidup , cerminnya jangan cermin angker dong , tidak bisa ngaca , ac nya dibenarkan dong biar tidak matimati sendiri , tolong seranggaserangga dibasmi juga , tetapi petugas dan staf ramah maksimal ! lokasi hotel dekat pvj banget , lokasi bagus sayang kalau tidak dimaksimalkan , terimakasih .	(petugas, ramah maksimal, positif) (staf, ramah maksimal, positif) (lokasi, bagus, positif)	(petugas, ramah maksimal, positif) (staf, ramah maksimal, positif) (lokasi, bagus, positif) (ac nya, matimati sendiri, negatif)

A.3 Penambahan Anotasi Pasangan Triplet Opini dengan Aspek Implisit

Terdapat 2181 pasangan triplet opini dengan aspek implisit yang ditambahkan dari ekspresi sentimen yang sebelumnya tidak memiliki pasangan ekspresi aspek eksplisit pada dataset. Contoh dari penambahan anotasi pasangan triplet opini eksplisit dapat dilihat pada Tabel A.3. Detail penambahan anotasi pasangan triplet opini eksplisit secara lebih lengkap dapat dilihat pada tautan berikut (https://docs.google.com/spreadsheets/d/1lrMzBeAfX7iMeFeHS307M35WRMLbtPTDtJ_4BXhWPTg/edit?usp=sharing).

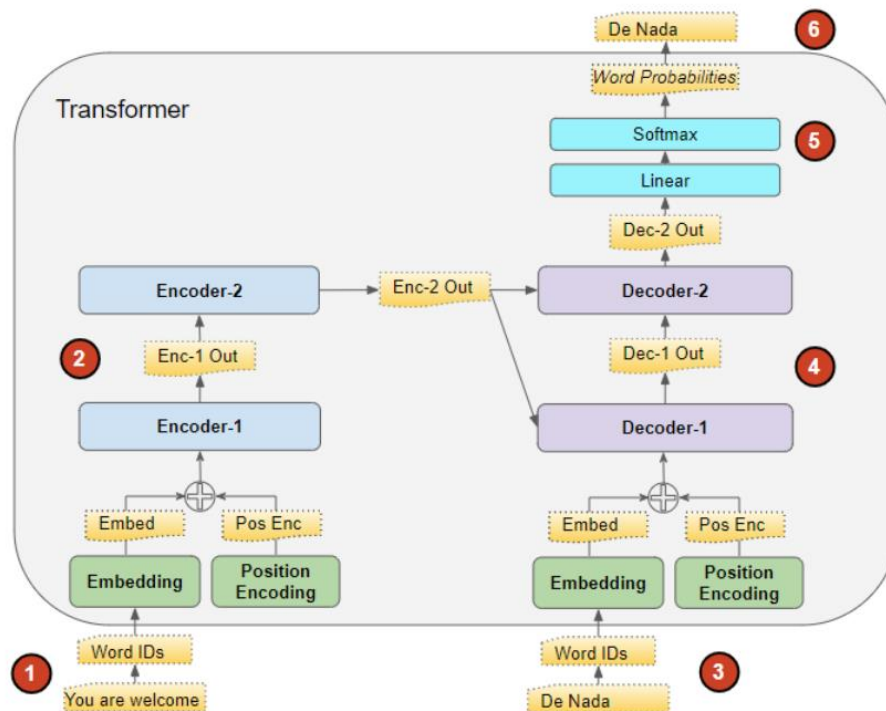
Tabel A.3 Contoh Penambahan Anotasi Pasangan Triplet Opini dengan Aspek Implisit

No.	Ulasan	Triplet Opini	Triplet eksplisit + Opini aspek implisit
1.	oke banget , tetapi ac nya tidak bisa diatur suhu nya .	(ac nya, tidak bisa diatur, negatif)	(ac nya, tidak bisa diatur, negatif) (_, oke banget, positif)
2.	bersih . tetapi sulit ditemukan .	-	(_, bersih, positif) (_, sulit ditemukan, negatif)
3.	bagus bersih nyaman , cuma tempat terlalu kedalam .	(tempat, terlalu kedalam, negatif)	(tempat, terlalu kedalam, negatif) (_, bagus, positif) (_, bersih, positif) (_, nyaman, positif)
4.	lumayan bersih tetapi lantainya tidak rata karena kurang semen di bawah nya , jadi menyebabkan pecah " .	(lantainya, tidak rata, negatif)	(lantainya, tidak rata, negatif) (_, lumayan bersih, positif)
5	recomended , hotel dengan price murah tetapi fasilitas oke punya . makanannya juga oke , pelayanan oke , bapak .	(price, murah, positif) (fasilitas, oke, positif) (makanannya, oke, positif) (pelayanan, oke, positif)	(price, murah, positif) (fasilitas, oke, positif) (makanannya, oke, positif) (pelayanan, oke, positif) (_, recommended, positif)

Lampiran B. Detail Cara Kerja Model Berbasis Transformer

B.1 Pelatihan Model

Pada bagian ini dijelaskan proses pelatihan model berbasis Transformer (Indo-T5). Ilustrasi pelatihan dapat dilihat pada Gambar B.1.



Gambar B.1 Ilustrasi Proses Pelatihan Transformer (Doshi, 2019)

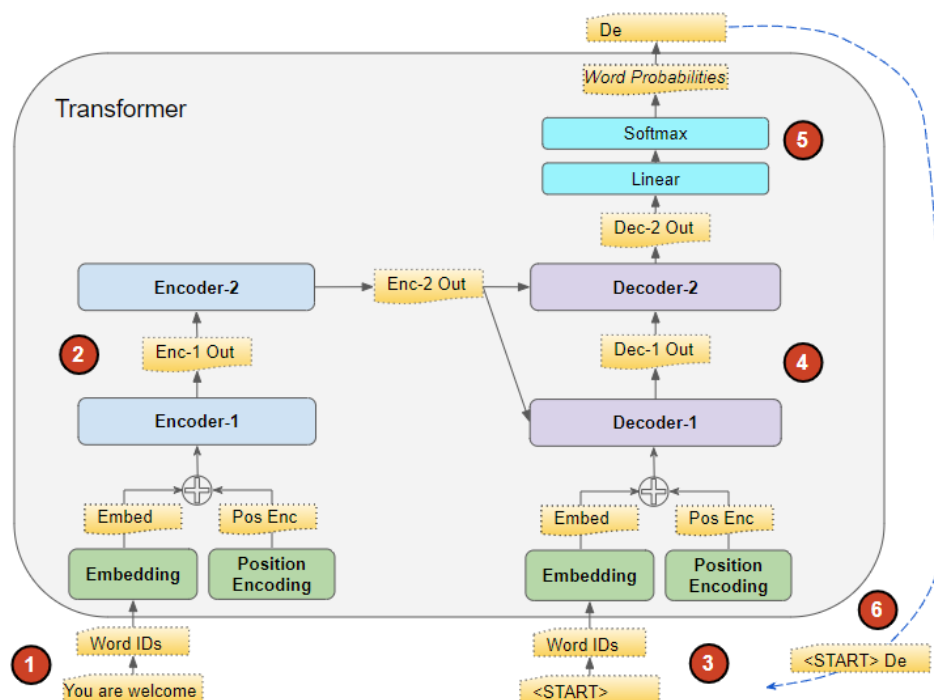
Proses pelatihan model berbasis transformer dapat dijabarkan sebagai berikut.

1. Masukan model berupa ID token dari teks diubah menjadi vektor representasi numerik teks oleh *layer embedding*. Selain itu, juga ditambahkan *positional encoding* (untuk memberikan informasi terkait posisi token).
2. Bagian *encoder* memproses vektor representasi numerik teks masukan pada tiap blok *encoder* untuk menghasilkan representasi baru yang akan digunakan oleh *decoder*.

3. Target dari teks masukan ditambahkan token spesial pada awal kalimat dan diubah menjadi menjadi vektor representasi numerik oleh layer *embedding* yang disertai dengan *positional encoding*.
4. Bagian *decoder* memproses target bersama dengan vektor dari *encoder* untuk menghasilkan representasi vektor untuk representasi urutan target.
5. Layer *output* mengubah vektor representasi urutan target menjadi probabilitas terhadap semua kata pada kumpulan kosa kata model.
6. Backpropagation dilakukan dengan *loss function* model transformer dengan membandingkan keluaran dengan target dari data pelatihan. Hal ini dilakukan untuk menghasilkan gradien untuk melatih model transformer untuk memperbarui bobot model.

B.2 Inferensi Model

Pada bagian ini dijelaskan proses inferensi model berbasis Transformer (Indo-T5). Ilustrasi pelatihan dapat dilihat pada Gambar B.2.



Gambar B.2 Ilustrasi Proses Inferensi Transformer (Doshi, 2019)

Terdapat perbedaan antara proses pelatihan model dan inferensi pada model berbasis Transformer. Tujuan inferensi model transformer adalah untuk menghasilkan keluaran berupa teks dari teks masukan saja.

Sama seperti model Seq2Seq pada umumnya, model berbasis transformer menghasilkan keluaran pada setiap *timestep* dan memasukkan keluaran dari *timestep* sebelumnya ke bagian *decoder* untuk *timestep* berikutnya sampai model membangkitkan token akhir kalimat. Namun, berbeda dengan model Seq2Seq, model berbasis transformer menggunakan seluruh keluaran yang telah dibangkitkan sampai *timestep* tersebut, bukan hanya keluaran pada *timestep* terakhir.

Proses inferensi model berbasis transformer dapat dijabarkan sebagai berikut.

1. Masukan model berupa ID token dari teks diubah menjadi vektor representasi numerik teks oleh *layer embedding*. Selain itu, juga ditambahkan *positional encoding* (untuk memberikan informasi terkait posisi token).
2. Bagian *encoder* memproses vektor representasi numerik teks masukan pada tiap blok *encoder* untuk menghasilkan representasi baru yang akan digunakan oleh *decoder*.
3. Berbeda dengan proses pelatihan, digunakan token spesial untuk mengawali proses inferensi pada bagian *decoder*.
4. Bagian *decoder* memproses target bersama dengan vektor dari bagian *encoder* untuk menghasilkan representasi vektor untuk urutan target.
5. Layer *output* mengubah vektor representasi urutan target menjadi probabilitas terhadap semua kata pada kumpulan kosa kata model.
6. Token terakhir dari keluaran layer output dipilih sebagai kata yang diprediksi. Token tersebut kemudian ditambahkan sebagai masukan dari layer decoder untuk *timestep* selanjutnya.
7. Langkah 1-6 diulang sampai model menghasilkan token spesial akhir kalimat (<eos>) atau keluaran telah melebihi panjang sekuens maksimum model.