

**PEMBELAJARAN TRANSFER DAN REPRESENTASI
BERBASIS SPAN UNTUK EKSTRAKSI TRIPLET OPINI
UNTUK ANALISIS SENTIMEN BERBASIS ASPEK**

TESIS

**Karya tulis sebagai salah satu syarat
untuk memperoleh gelar Magister dari
Institut Teknologi Bandung**

**Oleh
Rifo Ahmad Genadi
NIM: 23520033
(Program Studi Magister Informatika)**



**INSTITUT TEKNOLOGI BANDUNG
Januari 2021**

**PEMBELAJARAN TRANSFER DAN ` REPRESENTASI
BERBASIS SPAN UNTUK EKSTRAKSI TRIPLET OPINI
UNTUK ANALISIS SENTIMEN BERBASIS ASPEK**

Oleh
Rifo Ahmad Genadi
NIM: 23520033
(Program Studi Magister Informatika)

Institut Teknologi Bandung

Menyetujui
Tim Pembimbing

Tanggal 25 Januari 2020

Ketua

Masayu Leylia Khodra
NIP: 19760429 200812 2 001

DAFTAR ISI

HALAMAN PENGESAHAN.....	2
DAFTAR ISI	3
DAFTAR GAMBAR DAN ILUSTRASI	4
Bab I Pendahuluan	5
I.1 Latar Belakang.....	5
I.2 Masalah Penelitian.....	6
I.3 Tujuan	7
I.4 Hipotesis.....	7
I.5 Batasan Masalah.....	7
I.6 Metodologi.....	8
Bab II Tinjauan Pustaka	9
II.1 Pembelajaran Transfer di Pemrosesan Bahasa Alami	9
II.2 Mekanisme Atensi dan Transformers.....	10
II.3 BERT	11
II.4 XLM	13
II.5 Analisis Sentimen Berbasis Aspek.....	14
II.6 Ekstraksi Pasangan Ekspresi Aspek dan Opini dengan Representasi <i>Span</i>	15
II.7 Penelitian Terkait	16
II.7.1. Penelitian Zhao dkk. (2020).....	16
II.7.2. Tesis Azhar (2020)	18
Bab III Analisis Masalah dan Perancangan Solusi.....	20
III.1 Analisis Persoalan	20
III.2 Teknik Penyelesaian Masalah	21
III.3 Desain Eksperimen.....	23
III.4 Metode Evaluasi Kinerja	23
III.5 Rancangan Solusi	24
DAFTAR PUSTAKA	26

DAFTAR GAMBAR DAN ILUSTRASI

Gambar II.1. Perbedaan Proses pada Pembelajaran Mesin Tradisional dengan Pembelajaran Transfer. (Ruder, 2019)	9
Gambar II.2. Multi-head Attention (kiri) dan Arsitektur Model Transformers (kanan) (Vaswani dkk., 2017)	11
Gambar II.3. <i>Pre-training language model</i> lintas bahasa	13
Gambar II.4. Perbandingan Koekstraksi Ekspresi Aspek dan Ekspresi Opini dengan Ekstraksi Pasangan Ekspresi Aspek dan Ekspresi Opini (Zhao dkk., 2020)	16
Gambar II.5 <i>Framework SpanMlt</i> (Zhao dkk., 2020).....	18
Gambar II.6. Rincian Hasil Eksperimen Zhao dkk. (2020)	18
Gambar II.7. Arsitektur pembangunan model untuk sentence-pair-classification	19
Gambar III.1. Arsitektur Sistem Ekstraksi Triplet Opini Berbasis <i>Span</i>	25

Bab I Pendahuluan

I.1 Latar Belakang

Teks ulasan dapat ditelaah oleh organisasi atau pelaku bisnis untuk mengetahui pandangan konsumen terhadap suatu produk atau layanan. Hal tersebut juga dapat membantu organisasi untuk mengambil keputusan yang tepat dan relevan, serta mengetahui bagian yang perlu diperbaiki pada produk atau layanannya. Bidang studi yang menganalisis sentimen seseorang terhadap suatu entitas beserta atributnya dinamakan analisis sentimen atau penggalian opini (Liu, 2020). Analisis sentimen berbasis aspek (ASBA) dapat digunakan untuk hasil ekstraksi sentimen pada level yang rinci, yakni didapatkan target opini secara spesifik. ASBA dapat menghasilkan informasi sentimen terhadap aspek-aspek dari produk atau layanan yang terkandung dalam suatu opini. ASBA mencakup beberapa *subtask* dan penelitian-penelitian untuk ASBA biasanya mengerjakan beberapa *subtask* tersebut.

Penelitian perihal analisis sentimen berbasis aspek untuk domain hotel berbahasa Indonesia telah dilakukan oleh Genadi (2020) untuk melakukan ekstraksi triplet opini. Genadi (2020) mengadaptasi *framework* DOER (Luo dkk., 2019) untuk menyelesaikan keempat *subtask* tersebut, *framework* tersebut terdiri atas dua *stacked RNN* dan sebuah *cross-shared-unit* di antaranya. Konfigurasi terbaik pada penelitian tersebut mendapatkan F1-score sebesar 0.74 untuk *subtask* koekstraksi ekspresi aspek-ekspresi opini-polaritas sentimen, F1-score 0.9081 untuk *subtask* ekstraksi ekspresi aspek-ekspresi opini, dan *macro F1-score* sebesar 0.7266 untuk *subtask* klasifikasi polaritas sentimen. Berdasarkan hasil tersebut, bagian penanganan klasifikasi polaritas sentimen dan pemasangan ekspresi aspek dengan ekspresi opini pada sistem tersebut masih kurang baik jika dibandingkan dengan penanganan ekstraksi ekspresi aspek-ekspresi opininya. Selain itu, sistem tersebut juga mengalami kasus OOV, hal ini diduga dapat menyebabkan kesalahan prediksi pada *subtask* mana pun dan mengurangi kemampuan generalisasi model.

Subtask ekstraksi ekspresi aspek, ekspresi opini, klasifikasi polaritas sentimen dan pemasangan ekspresi aspek dengan ekspresi opini dapat dilakukan dengan mengubah persoalan *sequence tagging* menjadi permasalahan klasifikasi berbasis

span seperti pada penelitian Zhao dkk. (2020). *Framework* klasifikasi berbasis *span* yang dilakukan Zhao dkk. (2020) berhasil mendapatkan kinerja *state-of-the-art* untuk ekstraksi pasangan ekspresi aspek-ekspresi opini untuk data ulasan restoran dan data ulasan laptop Semeval 2014, Semeval 2015, Semeval 2016 (Fan dkk., 2019; Wang dkk., 2017; Wang dkk., 2016). Kinerjanya untuk *subtask* ekstraksi ekspresi aspek dan ekspresi sentimen pun melampaui seluruh model *baseline*.

Lalu, untuk permasalahan OOV dan meningkatkan generalisasi model dapat ditangani menggunakan pembelajaran transfer (*transfer learning*) (Peters dkk., 2018; Devlin dkk., 2019) dari suatu model bahasa *pre-trained*. Pendekatan demikian telah banyak digunakan untuk berbagai *task* pemrosesan bahasa alami dan terbukti menghasilkan kinerja yang lebih baik (Wang dkk., 2018; Wang dkk., 2019; Xu dkk., 2020; Wilie dkk., 2020). Penelitian Azhar (2020) pun mengujikan pendekatan tersebut terhadap data ulasan hotel dengan melakukan *sentence-pair classification* dan model yang dihasilkan berhasil mendapatkan kinerja *macro F1-score* 0.9304 untuk data uji ulasan hotel II (periode Maret 2019 - Juli 2019) dan tidak ada lagi kasus OOV. Selain itu, saat ini terdapat banyak varian model bahasa *pre-trained* untuk bahasa Indonesia, seperti BERT multilingual (Devlin dkk., 2019), XLM (Conneau dkk., 2019), IndoBERT (Wilie dkk., 2020), dll.

Penelitian tesis ini bertujuan mengadaptasi *framework* yang diusulkan oleh Zhao dkk. (2020) untuk menyelesaikan *task* ekstraksi ekspresi aspek, ekstraksi ekspresi opini, klasifikasi polaritas sentimen, dan pemasangan ekspresi aspek dengan ekspresi opini. Model modifikasi DOER pada penelitian Genadi (2020) akan dijadikan *baseline* untuk penelitian Tesis ini, sebagai pembanding untuk menguji efektifitas dari pendekatan yang diadaptasi untuk permasalahan-permasalahan yang telah disebutkan.

I.2 Masalah Penelitian

Rumusan masalah pada penelitian tesis ini adalah bagaimana cara mengadaptasi teknik penyelesaian masalah dengan mengadaptasi metode pembelajaran transfer dan teknik klasifikasi berbasis *span* yang diusulkan pada penelitian Zhao dkk. (2020) pada *task* ASBA untuk meningkatkan kinerja model dan memperbaiki penanganan OOV dari penelitian sebelumnya.

I.3 Tujuan

Tujuan dari penelitian tesis ini adalah menghasilkan model analisis sentimen berbasis aspek yang mencakup penanganan ekstraksi ekspresi aspek, ekstraksi ekspresi opini, klasifikasi polaritas sentimen, dan ekstraksi relasi ekspresi aspek dengan ekspresi opini, atau disebut ekstraksi triplet opini.

I.4 Hipotesis

Premis-1: Transformasi permasalahan *sequence labelling* menjadi klasifikasi berbasis *span* dapat melampaui kinerja pendekatan *sequence labelling* untuk menangani analisis sentimen berbasis aspek, yaitu untuk *task* ekstraksi pasangan AT dan OT (Zhao dkk., 2020).

Premis-2: Pemanfaatan metode model bahasa *pre-train* kontekstual (Peters dkk., 2018; Devlin dkk., 2019) terbukti efektif untuk berbagai *task* pemrosesan bahasa NLP, baik untuk bahasa *high-resource* (Wang dkk., 2018; Wang dkk., 2019; Xu dkk., 2020) maupun *low-resource* (Wilie dkk., 2020).

Berdasarkan pada premis-1 disusun hipotesis:

Hipotesis-1: *Task* ekstraksi triplet opini pada analisis sentimen berbasis aspek dapat lebih efektif dilakukan dengan mengubah persoalan tersebut menjadi persoalan klasifikasi berbasis *span*.

Berdasarkan pada premis-2 disusun hipotesis:

Hipotesis-2: Kinerja model ekstraksi triplet opini untuk analisis sentimen berbasis aspek dapat meningkat jika memanfaatkan model bahasa *pre-trained*.

I.5 Batasan Masalah

Batasan masalah dalam penelitian tesis ini adalah sebagai berikut.

- 1) Teks ulasan yang digunakan adalah teks ulasan berbahasa Indonesia pada domain hotel.
- 2) Model bahasa *pre-trained* yang digunakan adalah BERT multilingual, IndoBERT dan XLMR.

I.6 Metodologi

Metodologi penggerjaan penelitian tesis ini adalah sebagai berikut.

1) Analisis Persoalan dan Solusi

Pada tahap analisis persoalan yang dialami oleh penelitian Genadi (2020) dalam melakukan ekstraksi triplet opini, kemudian dilakukan studi literatur alternatif solusi untuk menangani persoalan-persoalan yang muncul.

2) Konstruksi Korpus

Pada tahap ini dilakukan konstruksi korpus yang digunakan untuk melatih model ekstraksi triplet opini. Korpus yang digunakan sama dengan penelitian Genadi (2020), dengan tambahan label relasi ekspresi aspek dengan ekspresi opini.

3) Desain Eksperimen

Pada tahap desain eksperimen akan dilakukan perancangan arsitektur pembangunan model serta menentukan tujuan, skenario, serta skema eksperimen yang akan dilakukan.

4) Implementasi

Pada tahap implementasi akan dilakukan pembangunan modul-modul yang dibutuhkan, termasuk modul bantuan untuk menjalankan skenario eksperimen yang telah didefinisikan.

5) Eksperimen

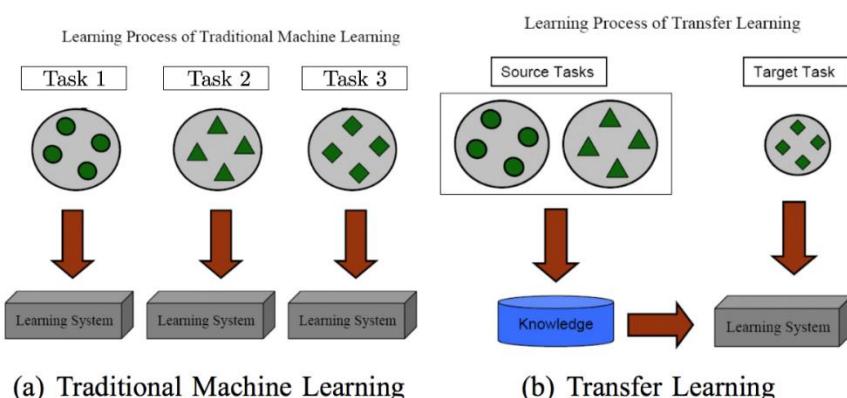
Pada tahap eksperimen akan dilakukan eksperimen untuk mengujikan efektifitas penggunaan klasifikasi berbasis *span* dan penggunaan model bahasa. Kemudian, akan dilakukan analisis terhadap hasil eksperimen.

Bab II Tinjauan Pustaka

II.1 Pembelajaran Transfer di Pemrosesan Bahasa Alami

Algoritma pembelajaran mesin mampu menyelesaikan berbagai jenis task pada pemrosesan bahasa alami dengan kinerja yang baik. Tetapi kinerja tersebut baru dapat tercapai apabila tersedia cukup banyak data yang spesifik untuk task tersebut dan teranotasi, atau dikenal dengan Teknik supervised learning. Pelatihan model pembelajaran mesin untuk dapat melakukan task spesifik biasanya dilakukan dari awal (Ruder, 2019). Hal tersebut menjadi masalah, karena data teranotasi tidak selalu tersedia dalam jumlah banyak, dan model selalu belajar dari nol, tanpa adanya pemanfaatan dari task atau domain yang terkait.

Teknik pembelajaran transfer merupakan salah satu pendekatan yang dapat mengatasi permasalahan tersebut. Teknik tersebut memungkinkan pengetahuan dari suatu task (source task) untuk meningkatkan kinerja model pada *task* yang dituju (target task). Hal ini intuitif dilakukan pada pemrosesan bahasa alami, karena nyatanya banyak *task* pada pemrosesan bahasa alami yang mengandung pengetahuan yang sama mengenai Bahasa (secara sintaksis ataupun semantik) dan dapat saling berbagi informasi akan hal tersebut. Perbandingan proses pembelajaran mesin tradisional dengan pembelajaran transfer dapat dilihat pada Gambar II.1.



Gambar II.1. Perbedaan Proses pada Pembelajaran Mesin Tradisional dengan Pembelajaran Transfer. (Ruder, 2019)

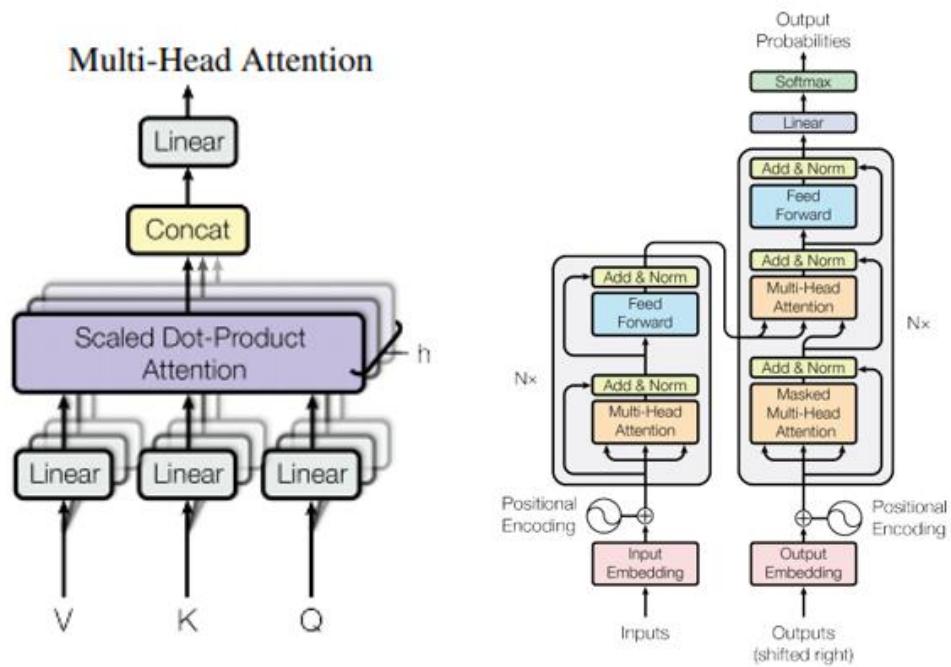
Formalnya, definisi dari pembelajaran transfer adalah sebagai berikut. Terdapat suatu domain asal DS, task pembelajaran TS, domain target DT, dan task pembelajaran TT, tujuan dari pembelajaran transfer adalah untuk meningkatkan pembelajaran fungsi target FT pada DT dengan memanfaatkan pengetahuan pada DS atau TS. (Pan & Yang, 2010).

Kemudian, di bidang pemrosesan bahasa alami, penggunaan model bahasa pre-trained kontekstual (Peters dkk., 2018; Devlin dkk., 2019) telah terbukti melampaui model-model state-of-the-art sebelumnya pada berbagai task di NLP (Wang dkk., 2018; Wang dkk., 2019; Xu dkk., 2020), termasuk pada bahasa low-resource seperti bahasa Indonesia (Wilie dkk., 2020). Model bahasa tersebut dilatih secara *unsupervised* pada data yang tak teranotasi berukuran besar (Devlin dkk., 2019).

II.2 Mekanisme Atensi dan Transformers

Bahdanau dkk. (2014) memperkenalkan mekanisme atensi untuk menangani persoalan pada LSTM, yaitu kesulitan untuk mempertahankan *long-term memory* saat kalimat yang diproses lebih panjang dibandingkan dengan kalimat pada data latihnya. Mekanisme atensi memungkinkan *decoder* untuk memberikan pembobotan terhadap bagian-bagian pada input, sehingga dapat dicari bagian mana saja dari input yang relevan untuk memprediksi target tertentu. Karenanya, pada arsitektur *encoder-decoder*, bagian *encoder* hanya memilih subset tertentu dari sekuens vektor yang dihasilkannya dari *input* (Bahdanau dkk., 2014).

Tetapi, masih ada kekurangan dari model rekuren seperti LSTM, yakni ia hanya mampu memproses *input* secara sekuensial per *step*-nya, sehingga tidak dimungkinkan pemanfaatan GPU untuk memproses *input* secara paralel (Vaswani dkk., 2017). Vaswani dkk. (2017) kemudian mengusulkan arsitektur Transformers yang memungkinkan pemrosesan keseluruhan sekuens input secara sekaligus. Komponen utama dari Transformers adalah *Multi-Head Attention* yang tersusun atas beberapa *attention layer* yang berjalan secara paralel, masing-masing layer tersebut mengangkap informasi mengenai berbagai aspek berbeda yang terdapat dalam sekuens input. Ilustrasi Multi-head Attention dan arsitektur Transformers dapat dilihat pada Gambar II.2.



Gambar II.2. Multi-head Attention (kiri) dan Arsitektur Model Transformers (kanan) (Vaswani dkk., 2017)

Bagian *encoder* dari Transformers tersusun dari enam *layer* identik, setiap *layer*-nya tersusun dari dua *sublayer* yaitu multi-head attention dan *feed forward neural network*. Bagian *decoder* pun terdiri dari enam *layer* identik, tetapi ia terdiri dari 3 *sublayer*-nya, selain dua *sublayer* seperti pada bagian *encoder*, pada bagian *decoder* terdapat tambahan *sublayer masked multi-head attention* untuk menangani *output* dari komponen *encoder*.

II.3 BERT

Bidirectional Encoder Representations from Transformers atau BERT adalah model bahasa yang memanfaatkan mekanisme *encoder* pada *Transformers*. BERT memroses sekuens kata untuk kedua arah, berbeda dengan teknik-teknik model bahasa sebelumnya yang membaca sekuens kata input secara sekuensial saja dalam satu arah (Devlin dkk., 2019). Pemrosesan secara dua arah tersebut, model bahasa yang dibentuk pun terbukti lebih baik dalam memahami konteks kalimat. BERT dapat dimanfaatkan untuk berbagai *task* pada NLP melalui proses adaptasi (*fine-tune*) dengan hanya menambahkan satu *layer* output. Saat ini terdapat banyak varian dan pengembangan dari BERT, perbedaan tiap varian dapat dilihat dari

ukuran modelnya, *cased/uncased*, bahasa korpus yang digunakan untuk pelatihannya, atau domain korpus yang digunakan untuk pelatihannya. Misalnya, terdapat model BERT-*base-uncased*, BERT-*large-cased*, Multilingual BERT, Distilbert (Sanh dkk., 2019), Roberta (Liu dkk., 2019), IndoBERT (Wilie dkk., 2020), dll.

Terdapat dua proses utama dalam BERT, yaitu tahap *pre-training* dan adaptasi (*fine-tuning*). Proses *pre-training* pada BERT menggunakan data tak beranotasi dalam jumlah besar, ada dua *task* yang dilakukan pada proses *pre-training* ini:

a) *Masked Language Model* (MLM)

BERT melakukan *masking* (mengganti token dengan token [MASK]) secara acak untuk beberapa token tertentu untuk melakukan task *language modeling* secara *bidirectional*. *Task* yang dilakukan adalah memprediksi kalimat asli yang utuh berdasarkan kalimat yang telah di-*masking* tersebut.

b) *Next Sentence Prediction* (NSP)

Model BERT dilatih terhadap *task* NSP karena beberapa *task* dalam NLP seperti NLI dan QA memerlukan informasi mengenai keterkaitan antara dua kalimat, yang mana hal tersebut tidak dapat diperoleh melalui pembelajaran MLM. *Task* yang dilakukan adalah menerima *input sentence pair* (kalimat A dan B), kemudian model memprediksi apakah kalimat B merupakan lanjutan dari kalimat A pada korpus.

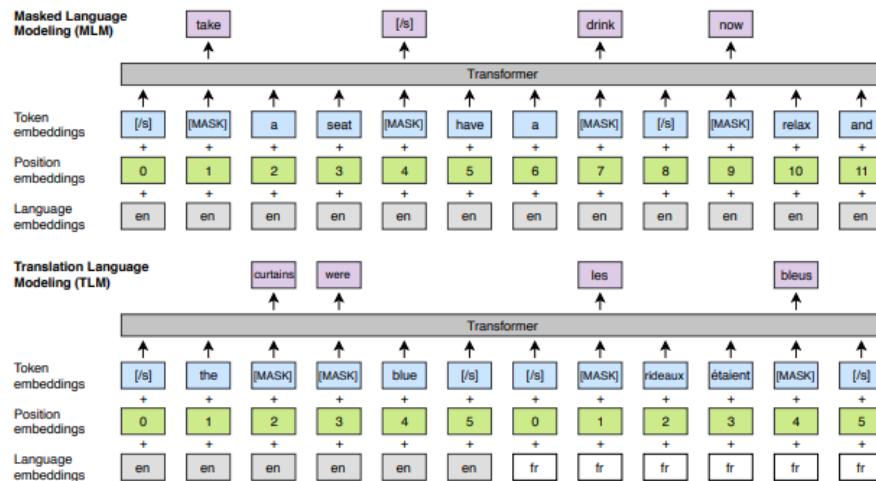
BERT menerima format masukan khusus dan perlu dilakukan praproses tersendiri. Secara umum, BERT melakukan tokenisasi terhadap level *sub-word*, serta menambahkan beberapa *token* khusus, seperti token [CLS] yang disematkan di awal kalimat dan token [SEP] yang menjadi penanda pemisah antar-kalimat. Setiap token ditambahkan *segment embeddings* untuk mengidentifikasi suatu token tergabung pada segmen kalimat mana, *token embedding* yang didapatkan dari proses tokenisasi biasa, dan *positional embeddings* yang menandakan posisi dari setiap *token* untuk keseluruhan sekvens input. Terakhir, representasi *input* dari BERT adalah penjumlahan dari *token embeddings*, *segment embeddings*, dan *positional embeddings*.

Kemudian, proses adaptasi BERT dilakukan agar BERT dapat digunakan terhadap *task* yang diharapkan. Hal ini cukup dilakukan dengan menambahkan satu *output*

layer (classification layer), biasanya berupa *feed forward neural network*, di atas *output layer* dari Transformers. Keluaran dari *output layer* tersebut akan menjadi *input* dari *classification layer* yang melakukan pembelajaran terhadap *downstream task* yang hendak dilakukan.

II.4 XLM

Conneau dkk. (2019) mengusulkan cara untuk mengefektifkan *pre-training* terhadap banyak dan lintas bahasa (*cross-lingual*). Pada penelitian tersebut diusulkan dua metode untuk mempelajari *cross-lingual language models* (XLMs): satu teknik *unsupervised* yang bergantung terhadap data monolingual, dan satu teknik *supervised* yang menggunakan data *cross-lingual* secara paralel. Selain MLM, pada XLM dilakukan *causal language modeling* (CLM) dan *translation language modeling* (TLM). Pada *task CLM*, model bahasa Transformers dilatih untuk memodelkan probabilitas kemunculan suatu kata, diberikan sekuens kata sebelumnya $P(w_t | w_1, \dots, w_{t-1}, \theta)$. Sedangkan, objektif dari TLM adalah ekstensi dari MLM, tetapi *input* yang digunakan adalah konkatenasi dari dua kalimat paralel, seperti yang diilustrasikan pada Gambar II.3. Sehingga, dalam memprediksi suatu kata berbahasa Inggris yang di-mask, model melakukan prediksi berdasarkan kata berbahasa Inggris di sekitarnya, atau berdasarkan terjemahan bahasa Perancis-nya.



Gambar II.3. *Pre-training language model* lintas bahasa

II.5 Analisis Sentimen Berbasis Aspek

Analisis sentimen berbasis aspek (ASBA) merupakan analisis sentimen yang dilakukan pada tingkat granularitas yang rinci, yakni hingga didapatkan sentimen untuk setiap aspeknya. Analisis sentimen berbasis aspek mencakup beberapa *subtask* dan dapat dilakukan dengan beberapa pendekatan (Pontiki dkk, 2016; Chen dkk., 2018). Pendekatan-pendekatan yang dilakukan biasanya mencakup beberapa *subtask*. Berikut merupakan penjelasan mengenai *subtask* dalam ASBA.

a) Kategorisasi Aspek

Pada kategorisasi aspek dilakukan pendekatan kategori-kategori yang disebutkan pada suatu kalimat. Kategori yang dideteksi adalah kategori yang didefinisikan di awal. Misalnya, jika didefinisikan kategori $C = \{\text{kebersihan}, \text{pelayanan}, \text{ac}, \text{internet}, \text{sunrise_meal}\}$, maka hasil kategorisasi aspek untuk ulasan “Ada beberapa fasilitas yg tidak diberikan yaitu sarapan dan wifi” adalah $\{\text{internet}, \text{sunrise_meal}\}$.

b) Ekstraksi Ekspresi Aspek / *Aspect Term Extraction* (ATE)

Pada ekstraksi ekspresi aspek dilakukan ekstraksi kata atau frasa yang merupakan aspek (ekspressi aspek). Ekspresi aspek yang diekstraksi merupakan bagian dari kalimat masukan. Misalnya, untuk ulasan “Ada beberapa fasilitas yg tidak diberikan yaitu sarapan dan wifi”, hasil ekstraksi ekspresi aspeknya adalah $\{\text{sarapan}, \text{wifi}\}$.

c) Ekstraksi Ekspresi Opini / *Opinion Term Extraction* (OTE)

Pada ekstraksi ekspresi opini dilakukan ekstraksi kata atau frasa yang menunjukkan sentimen dari penulis ulasan. Ekspresi opini yang diekstraksi merupakan bagian dari kalimat masukan. Misalnya, untuk ulasan “Ada beberapa fasilitas yg tidak diberikan yaitu sarapan dan wifi”, hasil ekstraksi ekspresi opininya adalah $\{\text{tidak diberikan}\}$.

d) Klasifikasi Polaritas Sentimen / *Aspect Sentiment Classification* (ASC)

Pada klasifikasi polaritas sentimen dilakukan klasifikasi setiap aspek (bisa kategori atau ekspresi aspek) menjadi positif, negatif, atau bisa juga ditambahkan label netral. Misalnya, untuk contoh pada poin a, polaritas sentimen untuk kategori internet dan sunrise_meal adalah negatif.

Lalu, untuk contoh pada poin b, polaritas sentimen untuk sarapan dan wifi adalah negatif.

e) Ekstraksi Relasi Ekspresi Aspek dan Ekspresi Opini

Pada task ekstraksi relasi ekspresi aspek dan ekspresi opini, dilakukan pemasangan antara hasil yang telah didapatkan dari ATE dan OTE. Misalnya, berdasarkan contoh sebelumnya, dari daftar ekspresi aspek yang telah diekstraksi {sarapan, wifi} dan daftar ekspresi opini {tidak diberikan}, hasil ekstraksi relasinya adalah {(sarapan, tidak diberikan), (wifi, tidak diberikan)}.

II.6 Ekstraksi Pasangan Ekspresi Aspek dan Opini dengan Representasi *Span*

PAOTE (*Pair-wise Aspect and Opinion Term Extraction*) merupakan *task* dalam analisis sentimen berbasis aspek yang diusulkan oleh Zhao dkk. (2020). Definisi formalnya, diberikan suatu kalimat $S = \{W_1, W_2, \dots, W_N\}$ yang terdiri dari N buah kata, *task* PAOTE mengekstraksi himpunan dari seluruh ekspresi aspek $AT = \{at_1, at_2, \dots, at_i\}$, himpunan dari seluruh ekspresi opini $OT = \{ot_1, ot_2, \dots, ot_j\}$, dan himpunan seluruh pasangan (AT, OT) $P = \{(at_m, ot_n), \dots\}$ yang terdapat pada ulasan. *Task* tersebut dilakukan dengan skema berbasis *span*. Mula-mula, seluruh *span* $SP = \{s_1, s_2, \dots, s_k\}$ yang mungkin dienumerasikan dari kalimat ulasan, masing-masing *span* tersebut adalah potongan (hingga panjang tertentu) dari kalimat masukan.

Berdasarkan kandidat *span-span* tersebut, keluarannya adalah 1) jenis ekspresi (ekspresi aspek atau ekspresi opini); 2) relasi R dari seluruh pasangan *span* $SP \times SP$ untuk mengidentifikasi pasangan (AT, OT) (Zhao dkk., 2020). Definisi formal dari kedua *subtask* tersebut adalah.

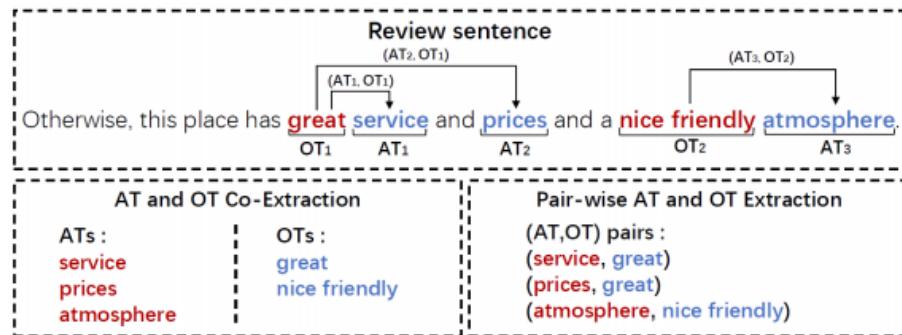
- a) Klasifikasi Ekspresi mengklasifikasikan setiap kandidat *span* s_c ke salah satu dari himpunan label $T \in \{A, O, null\}$, A menandakan bahwa $s_c \in AT$, O menandakan bahwa $s_c \in OT$ dan $null$ menandakan bahwa *span* tidak tergolong sebagai AT maupun OT .
- b) Identifikasi Relasi Pasangan mengklasifikasikan setiap pasangan *span* terurut (s_{c1}, s_{c2}) ke label biner $R \in \{True, False\}$. Relasi tersebut merupakan

relasi berarah, s_{c1} berupa AT dan s_{c2} berupa OT. Label *True* menandakan bahwa s_{c1} dan s_{c2} berasosiasi.

Penggunaan representasi *span* memiliki beberapa keunggulan sebagai berikut 1) menghindari propagasi *error* seperti yang terjadi pada penanganan ekstraksi ekspresi aspek dan ekspresi opini secara *pipeline* pada penelitian Fan dkk. (2019) (Zhao dkk., 2020) 2) memiliki ruang pencarian yang lebih kecil jika dibandingkan dengan pendekatan *sequence labeling* yang kompleksitasnya merupakan *power set* dari label terhadap panjang kalimat (Zhao dkk., 2020) 3) Menangani permasalahan ekstraksi relasi *one-to-many* atau *many-to-one* yang dialami pendekatan *sequence labeling* (Zhao dkk., 2020), karena klasifikasi dilakukan terhadap setiap kemungkinan pasangan *span*.

II.7 Penelitian Terkait

II.7.1. Penelitian Zhao dkk. (2020)



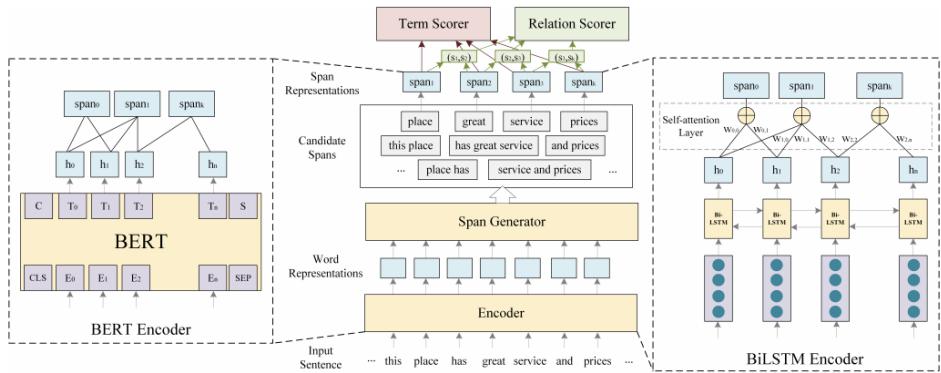
Gambar II.4. Perbandingan Koekstraksi Ekspresi Aspek dan Ekspresi Opini dengan Ekstraksi Pasangan Ekspresi Aspek dan Ekspresi Opini (Zhao dkk., 2020)

Ekstraksi ekspresi aspek dan ekspresi opini merupakan dua sub task dalam analisis sentimen berbasis aspek. Pasangan ekspresi aspek dan ekspresi sentimen dapat memberikan gambaran perihal opini terkait suatu produk atau layanan. Terdapat beberapa penelitian atau algoritma yang dapat melakukan koekstraksi kedua hal tersebut, seperti pada penelitian Wang dkk. (2016), Li & Lam (2017), Wang & Pan (2018), dan Yu dkk. (2019) namun metode-metode tersebut tidak menangani

pemasangan kedua jenis ekspresi tersebut (Zhao dkk., 2020). Misalkan, diberikan suatu kalimat ulasan seperti pada Gambar II.4, pasangan-pasangan ekspresi aspek-ekspresi opininya adalah (*service, great*), (*prices, great*), dan (*atmosphere, nice friendly*), tetapi hasil dari pendekatan koekstraksi pada penelitian-penelitian tersebut hanyalah himpunan masing-masing ekspresi aspek *{service, prices, atmosphere}* dan ekspresi opininya *{great, nice friendly}* saja (Zhao dkk., 2020). Zhao dkk. (2020) mengusulkan suatu *subtask* baru pada analisis sentimen berbasis aspek, yaitu PAOTE (*Pair-wise Aspect and Opinion Term Extraction*) yang detailnya telah dijelaskan pada subbab II.6. Perbedaan PAOTE dengan metode koekstraksi pada penelitian-penelitian sebelumnya yaitu *output* dari PAOTE yang berupa pasangan AT dan OT, berbeda dengan *task* koekstraksi yang *output*-nya berupa AT dan OT pada himpunan yang terpisah.

Penelitian tersebut juga mengusulkan suatu *multi-task framework* untuk menyelesaikan *task* PAOTE tersebut, yang dinamakan SpanMlt. Arsitektur dari *framework* tersebut dapat dilihat pada Gambar II.5. Arsitektur tersebut terdiri dari beberapa bagian sebagai berikut (Zhao dkk., 2020).

- a) *Base encoder* berupa BiLSTM atau BERT yang berfungsi untuk mengubah kalimat masukan menjadi representasi kata kontekstual.
- b) *Span generator* yang berfungsi untuk mengenumerasikan seluruh span yang mungkin dari suatu kalimat, yang direpresentasikan oleh *hidden output* dari *base encoder*.
- c) *Term scorer* yang berfungsi untuk mengklasifikasi span ke jenis ekspresi tertentu.
- d) *Relation scorer* yang berfungsi menentukan relasi korespondensi antara dua span.



Gambar II.5 *Framework SpanMlt* (Zhao dkk., 2020)

Varian-varian dari framework tersebut dievaluasikan terhadap beberapa dataset, yaitu data ulasan laptop dan restoran 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). Rincian hasil dari eksperimen tersebut dapat dilihat pada Gambar II.6. SpanMlt dengan setup model terbaik berhasil mendapatkan skor F1 yang lebih unggul pada seluruh dataset.

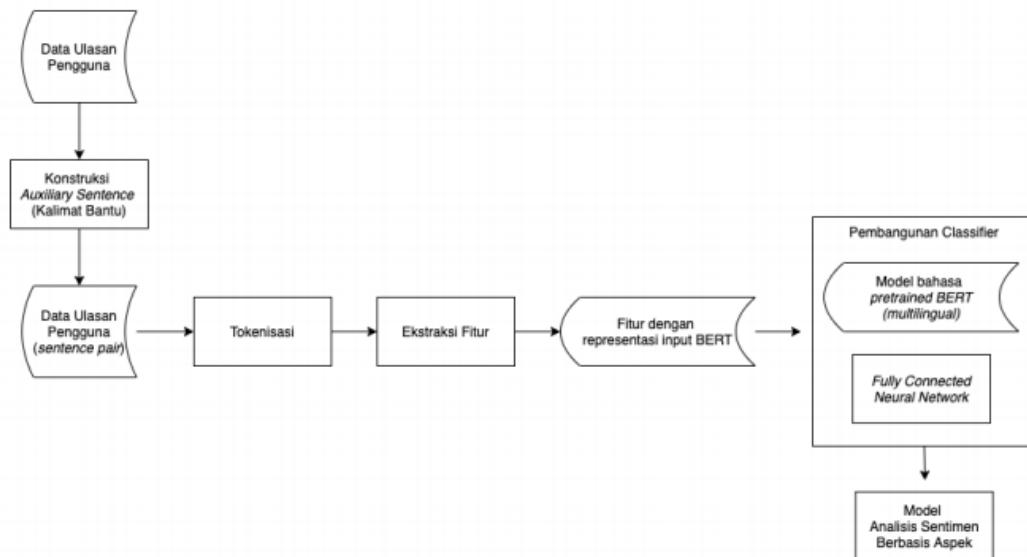
Models	14lap			14res			15res			16res		
	AT	OT	Pair									
BiLSTM+CRF	69.80	64.96	-	78.03	75.13	-	66.27	64.70	-	70.43	73.33	-
BERT+CRF	56.38	50.14	-	54.37	48.41	-	57.01	45.95	-	55.83	49.38	-
BERT+BiLSTM+CRF	56.99	51.33	-	54.08	51.53	-	55.85	47.79	-	55.18	51.53	-
RNCRF	74.92	67.21	-	75.18	67.95	-	74.14	64.50	-	73.12	65.51	-
CMLA	75.57	66.27	-	76.08	66.32	-	78.31	66.15	-	76.84	65.73	-
RNSCN	73.71	75.89	-	82.12	81.67	-	71.02	69.78	-	75.11	72.18	-
HAST+TOWE (pipeline)	79.14	67.50	53.41	82.56	75.10	62.39	79.84	68.45	58.12	81.44	75.71	63.84
JERE-MHS	74.61	64.02	52.34	79.79	77.44	66.02	75.00	71.38	59.64	76.08	78.02	67.65
SpanMlt (ours)	84.51	80.61	68.66	87.42	83.98	75.60	81.76	78.91	64.68	85.62	85.33	71.78

Gambar II.6. Rincian Hasil Eksperimen Zhao dkk. (2020)

II.7.2. Tesis Azhar (2020)

Azhar (2020) mengadaptasi penelitian Sun dkk. (2019) yang mentransformasikan persoalan *single sentence classification* menjadi *sentence-pair classification* untuk melakukan analisis sentimen berbasis aspek. Azhar (2020) menggunakan model bahasa *pre-trained* BERT Multilingual untuk mengubah kalimat masukan menjadi representasi teks kontekstual. Pada penelitian tersebut, model yang melakukan *sentence-pair classification* berhasil mendapatkan kinerja yang lebih baik dibandingkan model yang melakukan *single sentence classification*. Arsitektur

pembangunan model yang melakukan *sentence-pair classification* pada penelitian tersebut dapat dilihat pada Gambar II.7.



Gambar II.7. Arsitektur pembangunan model untuk sentence-pair-classification
(Azhar, 2020)

Hasil dari penelitian tersebut menunjukkan bahwa transformasi persoalan dapat membantu meningkatkan efektifitas pembelajaran transfer model *pre-trained* ke suatu *downstream task*. Selain itu, teknik penyelesaian masalah sebagai *sentence-pair classification* dari suatu model bahasa *pre-trained* dapat meningkatkan generalisasi model serta mengurangi jumlah kata-kata OOV untuk *task* analisis sentimen berbasis aspek berbahasa Indonesia pada domain hotel.

Bab III Analisis Masalah dan Perancangan Solusi

III.1 Analisis Persoalan

Berdasarkan evaluasi terhadap model pada penelitian sebelumnya serta studi terhadap literatur-literatur terbaru, masih terdapat beberapa permasalahan yang dapat ditangani. Permasalahan-permasalahan tersebut adalah sebagai berikut.

- a) Ruang pencarian yang besar

Ruang pencarian pada PAOTE cenderung lebih kecil jika dibandingkan dengan metode *sequence labelling*. Ruang pencarian pada *sequence labelling* merupakan *power set* dari seluruh kata pada ulasan.

- b) Inkonsistensi hasil prediksi polaritas sentimen pada satu ekspresi

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. Konsekuensinya, terdapat kemungkinan adanya perbedaan prediksi polaritas pada token yang seharusnya tergabung dalam satu ekspresi (Hu dkk., 2019). Misalnya, pada kalimat ulasan “Kamar oke. AC kurang dingin. Lampu kamar mandi ada yang mati.”, bisa jadi prediksi polaritas untuk ekspresi ‘lampa kamar mandi’ berbeda-beda untuk setiap tokennya. Persoalan tersebut tidak akan terjadi apabila klasifikasi polaritas sentimen dilakukan pada level *span*.

- c) Ekstraksi ekspresi aspek dan ekspresi sentimen secara terpisah

Metode *joint-tagging* menghasilkan himpunan ekspresi aspek dan ekspresi opini secara terpisah, seperti yang ditunjukkan pada Gambar II.3. Konsekuensinya, perlu ada modul tersendiri untuk menangani pemasangan antara ekspresi aspek dan ekspresi opini yang telah diekstraksi, hal ini dapat berupa model pembelajaran mesin tersendiri (Fan dkk., 2019) atau menggunakan metode *rule-based* seperti yang dilakukan pada penelitian sebelumnya, yaitu dengan memasangkan setiap aspek ke ekspresi opini yang terdekat. Metode *rule-based* tersebut memiliki keterbatasan dalam menangani relasi *one-to-many* atau *many-to-one* antara ekspresi aspek dan ekspresi opini pada kalimat ulasan. Misalnya, pada kalimat ulasan “Kamar luas dan kedap, bersih makanan enak. dan lokasi strategis.”, hasil yang

didapat dengan *rule* yang dibuat pada penelitian sebelumnya adalah untuk aspek makanan adalah (makanan, bersih) dan (makanan, enak). Lalu, jika salah satu dari hasil ekstraksi ekspresi aspek atau ekspresi opininya kosong, maka ia tidak akan menghasilkan triplet apa-apa.

- d) Kasus OOV. Berdasarkan pengujian pada penelitian sebelumnya didapatkan kasus OOV sebanyak 643 kata, 129 diantaranya merupakan bagian dari ekspresi aspek, 183 diantaranya merupakan bagian dari ekspresi opini, dan 331 diantaranya bukan bagian dari keduanya. Terdapat 45 ekspresi aspek dan 96 ekspresi opini yang gagal diekstraksi oleh model. Misalnya, pada kalimat ulasan “room bersih tetapi sayang linen kotor, bekas ada spot, kamar mandi bau, handuk sudah waktunya di ganti karena warnanya putih kehitaman, wifi susah (sinyal ada tetapi tidak mau terhubung)”, ekspresi opini ‘bekas ada spot’ dan ‘putih kehitaman’ tidak diekstraksi oleh model. Penelitian sebelumnya menggunakan model FastText untuk penanganan kasus OOV, tetapi representasi kata yang dihasilkan masih kurang baik. Misalnya, kata ‘*comfortable*’ muncul di data uji saja dan tidak muncul di data latih, saat dilihat *similarity*-nya dengan kata lain, kata-kata yang dianggap paling dekat adalah kata ‘mrawan, ‘ngrawan, dan ‘qayrawan’, berbeda dengan ekspektasi, misal ‘bahaya’ atau ‘rentan’. Penanganan kasus OOV ini juga dapat berpengaruh terhadap kemampuan generalisasi model.

III.2 Teknik Penyelesaian Masalah

Penelitian tesis ini akan mengadaptasi teknik penyelesaian masalah yang diusulkan oleh Zhao, dkk. (2020) yaitu melakukan klasifikasi berbasis *span* untuk mengekstraksi triplet opini, serta menggunakan *base encoder* berupa model bahasa *pre-trained* (Devlin dkk., 2019). Pendekatan ini dipilih karena dianggap dapat menyelesaikan persoalan-persoalan yang telah dijelaskan pada Bab III.1.

III.2.1. Klasifikasi dengan Representasi Berbasis *Span*

Penelitian Zhao dkk. (2020), seperti yang telah dijelaskan pada subbab II.7.1, berhasil menunjukkan efektifitas metode klasifikasi berbasis *span* dibandingkan

sequence labeling. Metode yang digunakan pada penelitian tesis ini mengadaptasi *framework* spanMLT yang telah dijelaskan di subbab II.7.1. *Framework* yang digunakan menyelesaikan *task* PAOTE yang telah dijelaskan pada subbab II.7, dengan perubahan pada bagian *relation scorer*-nya, bagian *relation scorer* pada *framework* tesis ini tidak hanya melakukan klasifikasi biner untuk menentukan valid tidaknya relasi antara *span* pertama dengan *span* yang kedua, melainkan melakukan klasifikasi multikelas terhadap masukan pasangan *span* (s_1, s_2) ke label $P \in \{PO, NG, O\}$, yang menunjukkan polaritas sentimen dari pasangan ekspresi aspek dan ekspresi opini. Pendekatan dan modifikasi tersebut dipilih karena dianggap dapat menyelesaikan ketiga permasalahan yang telah dijelaskan di subbab III.1.

Persoalan pertama, yakni mengenai ruang pencarian dapat teratasi dengan mentransformasikan persoalan *seqence labeling* ke klasifikasi berbasis *span*, ruang pencarian yang semula merupakan *power set* dari seluruh kata pada ulasan, menjadi bertambah secara linear saja dengan panjang kata (Hu dkk., 2019), sehingga mengurangi kompleksitas persoalan. Kemudian, persoalan inkonsistensi hasil klasifikasi polaritas sentimen dan ekstraksi ekspresi aspek-opini secara terpisah yang dialami pendekatan *sequence labeling* dapat dihindari dengan klasifikasi berbasis *span* karena *input* dari model adalah pasangan *span* dan akan menghasilkan satu buah polaritas sentimen saja terhadap pasangan *span* tersebut.

III.2.2. *Pre-trained Language Model* dan *Contextualized Word Embedding*

Seperti yang telah dijelaskan pada Bab II, pendekatan pembelajaran transfer dari suatu model bahasa *pre-trained* saat ini mampu memperoleh kinerja *state-of-the-art* pada berbagai *task* pemrosesan bahasa alami. Model bahasa yang mencapai kinerja *state-of-the-art* saat ini Sebagian besar menggunakan pendekatan *contextualized word embedding*, seperti BERT, XLM atau GPT.

Maka, pada penelitian tesis ini diujikan beberapa model bahasa yang dianggap sesuai untuk melakukan ekstraksi triplet opini bahasa Indonesia, yaitu IndoBERT (Wilie dkk., 2020) dan XLM-R (Conneau dkk., 2019). Model bahasa IndoBERT dipilih karena ia dilatih terhadap korpus Indo4B yang terdiri dari artikel berita, media social, Wikipedia, artikel daring, takarir rekaman video, serta dataset-dataset

bahasa Indonesia dan terbukti efektif untuk beberapa *benchmark*. XLM-R yang menggunakan teknik TLM pun terbukti efektif di beberapa *benchmark* untuk *task* berbahasa *low-resource*.

Pendekatan ini dipilih karena dianggap dapat meningkatkan kinerja model dan menangani kasus OOV, yaitu dengan teknik *wordpiece tokenization* yang dilakukan BERT. Lalu, penelitian ini menggunakan kedua model bahasa tersebut sebagai *base encoder* untuk *framework* spanMLT, selain itu diujikan juga efektifitas *pre-training* lanjutan menggunakan data korpus domain hotel berukuran 23 MB.

III.3 Desain Eksperimen

Terdapat dua eksperimen yang perlu dilakukan untuk membuktikan hipotesis penelitian tesis. Eksperimen pertama bertujuan untuk membandingkan metode *sequence labelling* dengan PAOTE yang merupakan metode ekstraksi berbasis *span*. Sementara itu, Eksperimen ketiga bertujuan menentukan kombinasi *hyperparameter* yang menghasilkan kinerja terbaik. *Hyperparameter* yang dipertimbangkan adalah *jenis pre-trained language model* yang digunakan, dilakukan atau tidaknya *pre-training* lebih lanjut terhadap data domain, lalu *batch size* dan *learning rate*-nya, seperti yang dilakukan pada penelitian Devlin dkk. (2019).

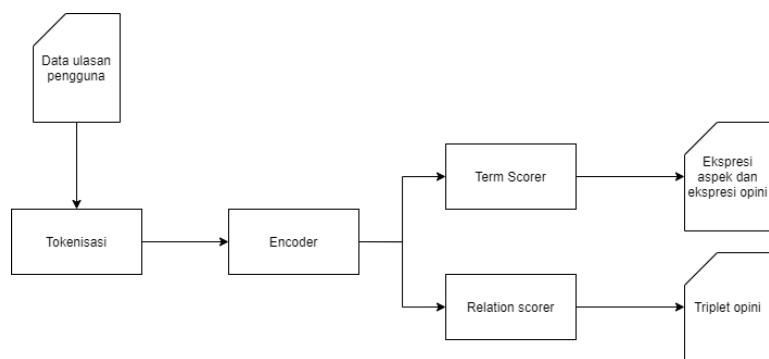
III.4 Metode Evaluasi Kinerja

Evaluasi kinerja dilakukan terhadap beberapa *subtask* pada analisis sentimen berbasis aspek: ekstraksi AT, ekstraksi OT, ekstraksi pasangan AT-OT, dan ekstraksi triplet opini. Model *baseline* yang digunakan sebagai pembanding adalah model pada penelitian Genadi (2020) untuk *subtask* ekstraksi AT, OT, dan ekstraksi triplet opini. Sedangkan, untuk *subtask* ekstraksi pasangan AT-OT, *framework* spanMLT (Zhao dkk., 2020) digunakan sebagai *baseline*. Metrik evaluasi kinerja yang digunakan adalah F1-score.

III.5 Rancangan Solusi

Secara umum, rancangan solusi yang dibuat mirip dengan *framework* pada penelitian Zhao dkk. (2020) yang telah dijelaskan pada subbab II.7.1. Tetapi, pada penelitian ini bagian *relation scorer* pada *framework* tersebut tidak hanya melakukan klasifikasi biner untuk menentukan valid tidaknya relasi antara *span* pertama dengan *span* yang kedua, melainkan melakukan klasifikasi multikelas terhadap masukan pasangan *span* (s_1, s_2) ke label $P \in \{\text{PO}, \text{NG}, \text{O}\}$, yang menunjukkan polaritas sentimen dari pasangan ekspresi aspek dan ekspresi opini. Arsitektur yang digunakan pada penelitian tesis ini dapat dilihat pada Gambar III.1. Mula-mula, data latih berupa ulasan hotel akan ditokenisasi, agar dapat diproses oleh model Transformers. Tokenizer akan melakukan *whitespace cleaning*, *punctuation splitting*, serta *wordpiece tokenization*. Contohnya, untuk kalimat “kamarnya bersih dan pelayanannya bagus”, akan ditokenisasi menjadi [“kama”, “##rny”, “##a”, “be”, “##rsi”, “##h”, “dan”, “pela”, “##yanan”, “##nya”, “bag”, “##us”].

Teks ulasan yang telah ditokenisasi kemudian diproses oleh *encoder* yang berupa model bahasa *pre-trained*. Keluaran dari *encoder* ini adalah representasi teks kontekstual yang kemudian menjadi masukan untuk klasifikasi yang dilakukan di modul selanjutnya. Terdapat dua *classifier* pada *framework* yang dibuat, yakni *term scorer* dan *relation scorer*. Kedua *classifier* tersebut berupa *fully connected neural network*. Perbedaan dari keduanya adalah masukan yang diterima, serta label keluarannya, karena tujuan dari keduanya pun berbeda, seperti yang telah dijelaskan pada subbab II.7.1. *Term scorer* menerima masukan berupa representasi teks untuk sebuah *span*, dan memiliki *node* sebanyak 3, yaitu $T \in \{\text{A}, \text{O}, \text{null}\}$. Sedangkan, *relation scorer* menerima masukan berupa representasi teks dari pasangan *span*, tepatnya $[p_i; p_j; p_i \odot p_j]$ yaitu hasil konkatenasi dari representasi teks *span* ke-i, representasi teks *span* ke-j, dan hasil perkalian elemen matriks dari keduanya. Implementasi rancangan solusi dilakukan menggunakan bahasa pemrograman Python, dengan library HuggingFace Transformers 4.2.2 (Wolf dkk., 2019), dengan *backend* Tensorflow 2 (Abadi dkk., 2016, Chollet dkk., 2018).



Gambar III.1. Arsitektur Sistem Ekstraksi Triplet Opini Berbasis *Span*

DAFTAR PUSTAKA

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265-283).
- Azhar, A. N. (2020). Pembelajaran Transfer untuk Analisis Sentimen Berbasis Aspek dalam Bahasa Indonesia menggunakan Pendekatan Fine-tuning Model Bahasa. *Graduate Thesis, Institut Teknologi Bandung, Bandung*
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- Chen, G., Zhang, Q., & Chen, D. (2018, June). A Pair-Wise Method for AspectBased Sentiment Analysis. In International Conference on Cognitive Computing (pp. 18-29). Springer, Cham.
- Chollet, F. (2018). Keras: The python deep learning library. *Astrophysics Source Code Library*, ascl-1806.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Fan, Z., Wu, Z., Dai, X., Huang, S., & Chen, J. (2019, June). Target-oriented opinion words extraction with target-fused neural sequence labeling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 2509-2518).
- Genadi, R. A. (2020). Koekstraksi Ekspresi dan Polaritas Untuk Analisis Sentimen Berbasis Aspek. *Undergraduate Thesis, Institut Teknologi Bandung, Bandung*
- Hu, M., Peng, Y., Huang, Z., Li, D., & Lv, Y. (2019). Open-domain targeted sentimen analysis via span-based extraction and classification. arXiv preprint arXiv:1906.03820.
- Li, X., & Lam, W. (2017, September). Deep multi-task learning for aspect term extraction with memory interaction. In Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 2886-2892).

- Liu, B. (2020). Sentimen analysis: Mining opinions, sentimens, and emotions. Cambridge university press.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Luo, H., Li, T., Liu, B., & Zhang, J. (2019). DOER: dual cross-shared RNN for aspect term-polarity co-extraction. *arXiv preprint arXiv:1906.01794*.
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 1345-1359.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A. S., ... & Hoste, V. (2016, June). SemEval-2016 task 5: ix Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30).
- Ruder, S. (2019). Neural Transfer Learning for Natural Language Processing. PhD Thesis, National University of Ireland, Galway.
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. *Proceedings of NAACLHLT*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.
- Wang, W., Pan, S. J., Dahlmeier, D., & Xiao, X. (2017, February). Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).
- Wang, W., Pan, S. J., Dahlmeier, D., & Xiao, X. (2016). Recursive neural conditional random fields for aspect-based sentimen analysis. *arXiv preprint arXiv:1603.06679*.
- Wang, W., & Pan, S. J. (2018, July). Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In Proceedings

of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 2171-2181).

Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., ... & Bowman, S. R. (2019). SuperGlue: A stickier benchmark for general-purpose language understanding systems. *arXiv preprint arXiv:1905.00537*.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*.

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., ... & Purwarianti, A. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. *arXiv preprint arXiv:2009.05387*.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2019). HuggingFace's Transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

Xu, L., Zhang, X., Li, L., Hu, H., Cao, C., Liu, W., ... & Lan, Z. (2020). Clue: A chinese language understanding evaluation benchmark. *arXiv preprint arXiv:2004.05986*.

Yu, J., Jiang, J., & Xia, R. (2018). Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(1), 168-177.

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).