

BAB 2 LANDASAN TEORI

Tinjauan teori ini digunakan sebagai dasar teori untuk menganalisa masalah dan perancangan solusi dalam penulisan laporan ini. Secara keseluruhan teori yang digunakan adalah kata konjungsi, *Natural Language Processing*, *Semantic Analysis*, model BERT, *Confusion Matrix*, dan penelitian yang sudah dilakukan sebelumnya.

2.1 Kata Konjungsi

Berdasarkan Kamus Besar Bahasa Indonesia versi enam daring, konjungsi adalah kata atau ungkapan penghubung antarkata, antarfrasa, antarklausa, dan antarkalimat [19]. Kata konjungsi diperlukan agar tidak terjadi kesalahpahaman antara satu klausa atau kalimat dengan klausa atau kalimat yang lain karena konjungsi berfungsi agar kalimat-kalimat tersebut saling berhubungan atau terhubung [26]. Berdasarkan perilaku sintaksisnya dalam kalimat, konjungsi atau kata hubung dapat dibagi menjadi empat jenis, konjungsi koordinatif, konjungsi subordinatif, konjungsi korelatif, dan konjungsi antarkalimat [27]. Selain itu, kata konjungsi dapat dibagi kembali berdasarkan kegunaannya seperti menunjukkan keterangan waktu, tujuan, sebab, atau hasil.

Penjelasan tiap jenis kata konjungsi menurut Moeliono (2017) dalam buku Tata Bahasa Baku Bahasa Indonesia Edisi keempat adalah sebagai berikut.

2.1.1 Konjungsi Koordinatif

Konjungsi koordinatif adalah konjungsi yang menghubungkan dua unsur atau lebih yang sama penting atau memiliki status sintaktis yang sama seperti dinyatakan di atasnya. Meskipun menggabungkan unsur yang sama setiap kata konjungsi koordinatif memiliki fungsi yang berbeda dalam penggunaannya. Contohnya adalah kata 'dan' menandakan hubungan penambahan dan kata 'atau' menandakan hubungan pemilihan. Beberapa kata konjungsi yang termasuk dalam jenis konjungsi koordinatif adalah *dan*, *atau*, *melainkan*, *padahal*, *sedangkan*, *serta*, *tetapi*, dan *dan/atau*.

2.1.2 Konjungsi Subordinatif

Konjungsi subordinatif adalah konjungsi yang menghubungkan dua klausa atau lebih dan klausa itu tidak memiliki status sintaktis yang sama. Dari perilaku sintaktis dan semantisnya, konjungsi subordinatif dapat dibagi menjadi tiga belas kelompok. Tiga belas kelompok tersebut adalah waktu, syarat, pengandaian, tujuan, konsesif, perbandingan, sebab, hasil, alat, cara, komplementasi, atributif, dan perbandingan. Contoh kata konjungsi subordinatif berdasarkan kelasnya adalah sebagai berikut.

1. Konjungsi Subordinatif Waktu

- menunjukkan awal peristiwa
 - sejak
 - sedari
 - semenjak
- menunjukkan awal suatu peristiwa yang ditandai dengan peristiwa lain
 - begitu
 - demi
 - ketika
 - sambil
 - selagi
- menunjukkan awal suatu peristiwa yang didahului dengan peristiwa lain
 - setelah
 - sebelum
 - sehabis
- menunjukkan lamanya suatu peristiwa yang ditandai dengan peristiwa tertentu
 - hingga
 - sampai

2. Konjungsi Subordinatif Syarat

- jikalau
- kalau

- manakala

3. Konjungsi Subordinatif Pengandaian

- andaikan
- seandainya
- sekiranya

4. Konjungsi Subordinatif Tujuan

- agar
- biar
- supaya

5. Konjungsi Subordinatif Konesif

- biarpun
- kendati(pun)
- meski(pun)

6. Konjungsi Subordinatif Perbandingan

- alih-alih
- daripada
- ibarat
- laksana

7. Konjungsi Subordinatif Sebab

- karena
- sebab
- oleh karena
- oleh sebab

8. Konjungsi Subordinatif Hasil

- maka(nya)
- sehingga

- sampai(-sampai)

9. Konjungsi Subordinatif Alat

- dengan
- tanpa

10. Konjungsi Subordinatif Cara

- dengan
- tanpa

11. Konjungsi Subordinatif Komplementasi

- bahwa

12. Konjungsi Subordinatif Atributif

- yang

13. Konjungsi Subordinatif Perbandingan

- ... sama ... dengan ...
- ... lebih ... dari ... dari(pada) ...

2.1.3 Konjungsi Korelatif

Konjungsi korelatif adalah sepasang konjungsi koordinatif yang menghubungkan dua kata, frasa, atau klausa yang memiliki status sintaktis yang sama. Konjungsi korelatif terdiri atas dua bagian yang umumnya terpisah satu dengan yang lain. Beberapa kata konjungsi yang termasuk dalam jenis konjungsi korelatif adalah *baik ... maupun ..., tidak ..., tetapi ..., dan betapapun ..., masih lebih ...*

2.1.4 Konjungsi Antarkalimat

Konjungsi antarkalimat menghubungkan satu kalimat dengan kalimat yang lain. Sedangkan konjungsi intrakalimat berbeda dengan konjungsi antarkalimat. Oleh karena itu, konjungsi antarkalimat selalu digunakan di awal kalimat baru yang huruf pertamanya dituliskan dengan huruf kapital dan di belakang

konjungsi tersebut diikuti tanda koma. Berikut adalah beberapa contoh konjungsi antarkalimat.

- Biarpun demikian, ...
- Sekalipun begitu, ...
- Walaupun demikian, ...
- Meskipun begitu, ...
- Sungguhpun demikian, ...
- Kemudian, ...
- Sesudah itu, ...
- Selanjutnya, ...

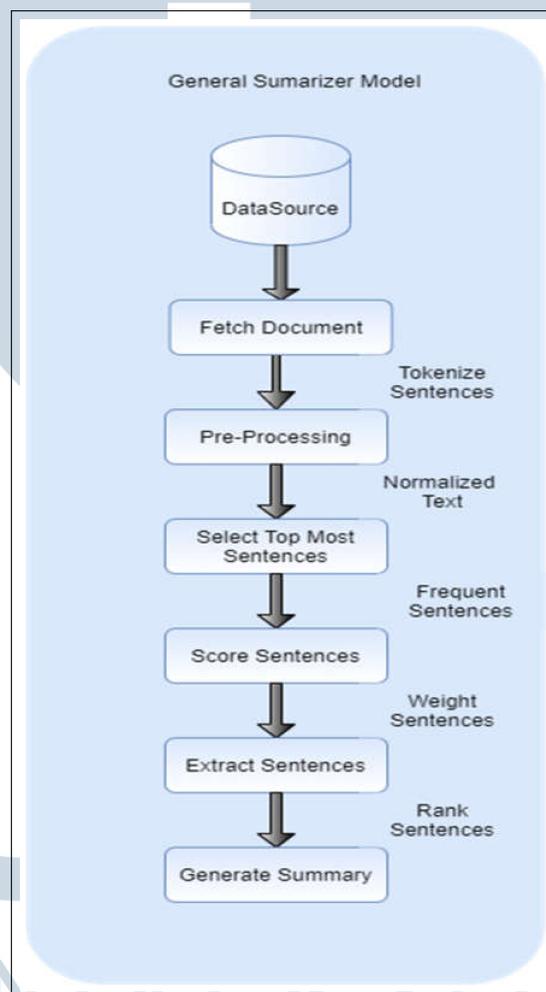
2.2 Natural Language Processing

Natural Language Processing (NLP) adalah kumpulan dari berbagai macam teknik komputasi untuk analisis secara otomatis dan representasi atas bahasa manusia yang termotivasi oleh teori [28]. Pada awal perkembangan NLP, pada sekitar akhir tahun 1940-an hingga akhir tahun 1960-an, NLP lebih berfokus pada representasi sintaksis atau struktur sintaksis dari teks dalam proses analisisnya seperti frekuensi penggunaan sebuah kata karena kebutuhan yang ada pada masa tersebut [29]. Tetapi algoritma memiliki batasan dalam pemahaman bahasa karena kemampuannya yang dapat melakukan analisa informasi hanya berdasarkan teks yang ada dan tidak dapat memproses informasi yang terkandung dalam teks tersebut [28]. Kemudian perkembangan NLP berpindah fokus pada analisis secara semantik untuk memenuhi kebutuhan pada masa itu untuk memproses teks yang memiliki arti ganda atau ambigu.

2.2.1 Semantic Analysis

Dalam linguistik umum, *Semantic Analysis* atau analisis semantik adalah kegiatan menganalisa makna kata, ekspresi, kalimat secara utuh, dan ucapan dalam konteks. Sedangkan dalam prakteknya, *semantic analysis* berarti menterjemahkan suatu ekspresi menjadi sebuah metabahasa semantik [30].

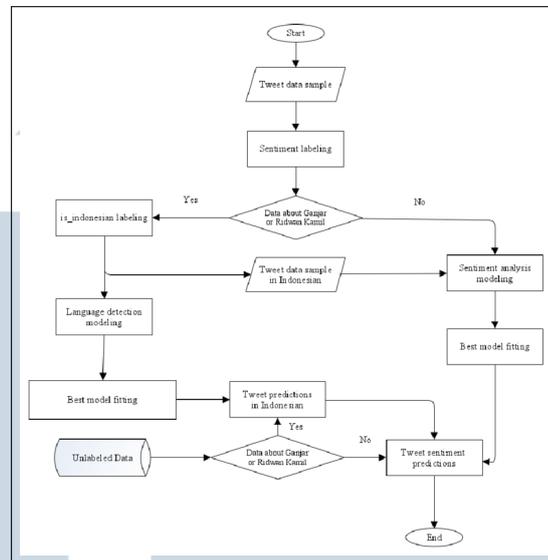
Umumnya, analisis semantik dilakukan dengan memproses teks terlebih dahulu yang terbagi ke dalam beberapa tahap. Pertama, menghapus *stop words* atau kata-kata yang sering muncul tetapi tidak memberikan informasi yang membantu. Setelah itu, melakukan proses *stemming*, proses untuk mengubah sebuah kata menjadi kata dasar. Terakhir, menggunakan teks yang sudah diproses ke dalam algoritma untuk melakukan analisis semantik [31]. Alur proses *Semantic Analysis* secara umum dapat dilihat pada Gambar 2.1



Gambar 2.1. Alur proses *Semantic Analysis* secara umum [1]

Tetapi proses yang dilalui untuk melakukan *Semantic Analysis* tidak akan sama untuk setiap kasus. Hal ini dikarenakan tujuan utama dari *Semantic Analysis* untuk melakukan ekstraksi arti dari kata atau kalimat. Maka alur proses *Semantic Analysis* akan berbeda untuk setiap implementasinya. Perbedaan tersebut dapat dilihat antara Gambar 2.1 dengan Gambar 2.2 yang merupakan alur proses *Semantic*

Analysis untuk Bahasa Inggris.

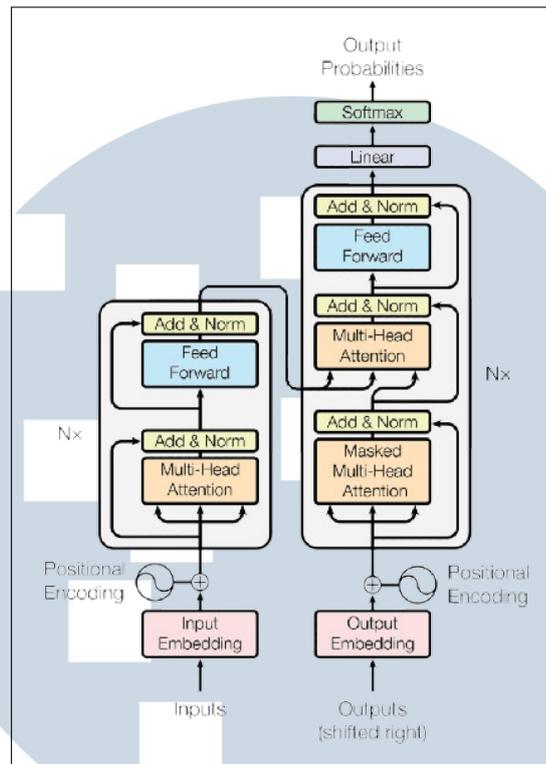


Gambar 2.2. Alur proses *Semantic Analysis* untuk Bahasa Inggris [2]

2.2.2 Model Transformer

Transformer adalah sebuah model berbentuk transduksi sekuensial menggunakan *multi-headed self-attention* [3]. Arsitektur Transformer memiliki struktur *encoder-decoder* dan menggunakan fungsi *self-attention* secara bertumpuk. Bentuk arsitektur Transformer dapat dilihat pada Gambar 2.3.

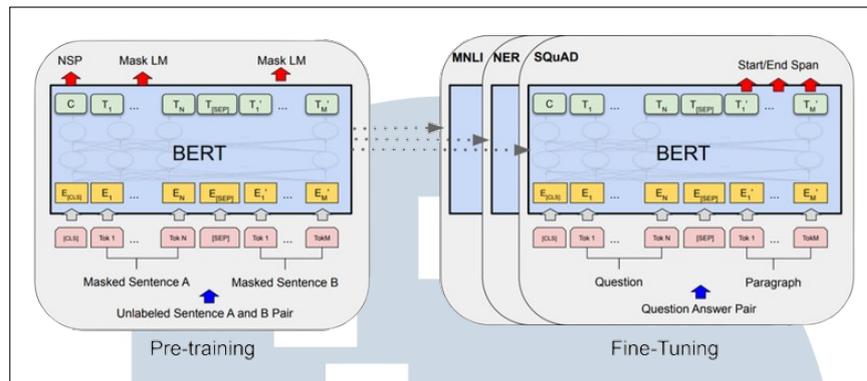




Gambar 2.3. Model arsitektur Transformer [3]

A Model BERT (*Bidirectional Encoder Representations from Transformers*)

BERT adalah salah satu model yang menggunakan arsitektur Transformer. Model ini dikembangkan oleh Google pada tahun 2018 yang dirancang untuk melakukan pra-pelatihan representasi dari teks yang tidak diberikan label dengan menggabungkan konteks dari kiri dan kanan di semua lapisan. Model arsitektur yang digunakan untuk model BERT tidak berubah jauh dari arsitektur Transformer, tetapi model BERT menggunakan dua tahap untuk melatih model, tahap *pre-training* dan tahap *fine-tuning*. Contoh *training* model dapat dilihat pada Gambar 2.4 [4]. Model BERT dapat mencapai nilai GLUE sebesar 80.5%, nilai akurasi MultiNLI sebesar 86.7%, dan uji F1 SQuAD v1.1 dan SQuAD v2.0 berturut-turut sebesar 93.2 dan 83.1 [4].



Gambar 2.4. Prosedur *pre-training* dan *fine-tuning* untuk BERT [4]

2.3 Confusion Matrix

Confusion Matrix (CM) adalah salah satu cara untuk menganalisa hasil sebuah model dengan menghitung jumlah prediksi benar dan salah lalu memasukkannya ke dalam sebuah matriks 2x2. Matriks tersebut kemudian dapat digunakan untuk menghitung *precision*, *recall*, dan *F1-score* untuk setiap hasil klasifikasi. Bentuk table CM dapat dilihat pada Tabel 2.1.

Tabel 2.1. *Confusion Matrix*

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Actual Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Setelah didapatkan CM, perhitungan *precision and recall* dan *F1-score* dapat dilakukan dengan rumus sebagai berikut [32].

$$Precision = \frac{TP}{TP + FP} \quad (2.1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.2)$$

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2.3)$$