

## BAB II

### LANDASAN TEORI

#### 2.1 N-gram Model

*N-gram* merupakan suatu model probabilitas yang memberikan prediksi kemunculan kata berikutnya berdasarkan probabilitas kemunculan kata  $N-1$  sebelumnya. *N-gram* digunakan untuk mengolah teks yang disebut *training corpus* kata setelah kata tertentu. Namun, jika kemunculan sebuah kata tersebut tidak ada, maka frekuensi kemunculan kata tersebut adalah nol (Prayogo, 2018). Rahmawan (2011) menambahkan bahwa *n-gram* dapat dibagi menjadi 2 kategori, yaitu berbasis kata dan berbasis karakter.

*N-gram* berbasis karakter merupakan potongan  $n$ -karakter yang diambil dari suatu *string*. Model *n-gram* digunakan untuk mengambil potongan-potongan dari karakter huruf sejumlah  $n$  dalam sebuah kata yang secara kontinu dari sumber hingga akhir dari *string* (Vernanda, 2018). Sebagai contoh, terdapat kata “COMPUTER” maka akan didapatkan *n-gram* sebagai berikut.

Unigram : C, O, M, P, U, T, E, R

Bigram : CO, OM, MP, PU, UT, TE, ER

Trigram : COM, OMP, MPU, PUT, UTE, TER

Quadgram : COMP, OMPU, MPUT, PUTE, UTER

Sedangkan *n-gram* berbasis kata akan memberikan potongan-potongan kata dari suatu *corpus* data. Sebagai contoh terdapat sebuah kalimat “He said bye as he walked through the door” maka akan didapatkan *n-gram* sebagai berikut (Kumar,

2017).

Unigram : He, said, bye, as, he, walked, through, the, door

Bigram : He said, said bye, bye as, as he, he walked, walked through, through the, the door

Trigram : He said bye, said bye as, bye as he, as he walked, he walked through, walked through the, through the door

Quadgram : He said bye as, said bye as he, bye as he walked, as he walked through, he walked through the, walked through the door

Rangkaian kata tersebut kemudian disimpan sebagai statistik pada *language model* yang didasarkan pada urutan kata dan kata yang paling sering digunakan dalam teks tanpa memperdulikan konteks dan *semantic affinity* yang didasarkan pada kemungkinan kata tersebut muncul bersama dalam urutan kalimat (Prayogo, 2018). Salah satu cara untuk membuat estimasi dari probabilitas ini adalah dari perhitungan frekuensi secara relatif dengan cara membuat sebuah *corpus* yang sangat besar (Daniel dkk., 2019).

## 2.2 Grammar Checker

*Grammar checker* adalah suatu aplikasi yang dapat melakukan pendeteksian atas kesalahan yang terjadi saat penulisan atau penggunaan bahasa. Proses dari sebuah aplikasi *grammar correction* mempunyai 4 langkah kerja yaitu, (Hariatsyah, 2013)

1. Pengidentifikasian bagian-bagian yang mempunyai kesalahan *grammar*.
2. Pengidentifikasian bagian-bagian yang melanggar aturan.
3. Pengidentifikasian kemungkinan penyebab kesalahan.

4. Pembuatan dan penyusunan kemungkinan perbaikan yang dapat dilakukan.

Terdapat 2 metode populer yang dapat mendeteksi kesalahan tata bahasa dalam sebuah kalimat. Metode pertama adalah menghasilkan *parse tree* secara lengkap dari sebuah kalimat untuk mengidentifikasi kesalahan dan metode yang kedua adalah berdasarkan aturan pengecekan yang mendeteksi urutan dalam sebuah teks (Konchady, 2009).

### **2.3 Bahasa Indonesia**

Bahasa Indonesia lahir pada tanggal 28 Oktober 1928 saat para pemuda dari berbagai pelosok Nusantara berkumpul dalam kerapatan Pemuda dan berikrar (1) bertumpah darah yang satu, tanah Indonesia, (2) berbangsa yang satu, bangsa Indonesia, dan (3) menjunjung bahasa persatuan, bahasa Indonesia yang dikenal dengan Sumpah Pemuda. Pada tahun 1928 itulah bahasa Indonesia dinyatakan kedudukannya sebagai bahasa negara pada tanggal 18 Agustus 1945. Kemudian, pada tahun 1954 dinyatakan bahwa bahasa Indonesia berasal dari bahasa Melayu yang tumbuh dan berkembang sejak zaman dulu yang digunakan sebagai bahasa perhubungan di seluruh Asia Tenggara (Kemendikbud, 2019).

Pada tanggal 12 Oktober 1972, Menteri Pendidikan dan Kebudayaan kemudian menyusun pedoman umum yang merupakan pemaparan ejaan. Pedoman yang dibuat ini berisi pedoman untuk penggunaan huruf, pemenggalan kata, pemakaian huruf kapital dan huruf miring, penulisan kata, singkatan dan akronim, penulisan unsur serapan dan pemakaian tanda baca (Hariatsyah, 2013)

#### **2.3.1 Tata Bahasa**

Tata bahasa merupakan ilmu yang mempelajari kaidah yang mengatur

penggunaan bahasa. Tata bahasa meliputi ilmu tata bentuk kata (morfologi) dan ilmu tata kalimat (sintaksis). Struktur ini kalimat bahasa Indonesia ragam tulis sebenarnya sangat sederhana, yaitu hanya berupa subjek dan predikat (S-P). Struktur inti tersebut dapat diperluas menjadi beberapa tipe kalimat dasar. Contoh sebagai berikut ini (badanbahasa.kemendikbud.go.id, 2020).

- a. *Anak itu sering melamun.* (Subyek + Predikat)
- b. *Sukarno dan Mohammad Hatta mempersatukan bangsa ini.* (Subyek + Predikat + Obyek)
- c. *Ajaran Mahatama Gandhi ditakuti penjajah Inggris.* (Subyek + Predikat + Pelengkap)
- d. *Raja Jawa menghadiahi VOC Pesisir Utara Pulau Jawa.* (Subyek + Predikat + Obyek + Pelengkap)
- e. *Jamu itu sangat baik untuk kesehatan.* (Subyek + Predikat + Keterangan)
- f. *Zulkarnain membersihkan tinta itu dengan sabun.* (Subyek + Predikat + Obyek + Keterangan)

Berdasarkan beberapa contoh di atas, tampak bahwa struktur inti kalimat bahasa Indonesia adalah *subyek + predikat* yang dapat ditambah dengan *obyek, pelengkap, dan/atau keterangan*  $S + P + (\{O\} + \{Pel\} + \{K\})$ . Struktur inti kalimat tersebut dapat diperluas menjadi beberapa tipe kalimat dasar. Yang dimaksud dengan kalimat dasar adalah kalimat yang terdiri atas satu klausa, lengkap unsur-unsurnya, dan paling lazim pola urutannya. Struktur kalimat dasar bahasa Indonesia dapat dikelompokkan ke dalam beberapa tipe berikut (badanbahasa.kemendikbud.go.id, 2020).

- (1) subyek – predikat (S-P)

- (2) subyek – predikat – obyek (S-P-O)
- (3) subyek – predikat – pelengkap (S-P-Pel)
- (4) subyek – predikat – obyek – pelengkap (S-P-O-Pel)
- (5) subyek – predikat – obyek – keterangan (S-P-O-K)
- (6) subyek – predikat – keterangan (S-P-K)

a. Ciri-ciri Subyek

Subyek merupakan salah satu fungsi dalam kalimat yang merupakan bagian klausa yang menjadi pokok kalimat. Subyek dapat berupa kata benda (*nomina*), kelompok kata benda (*frasa nominal*), atau klausa. Subyek dapat pula disertai kata *itu*. Subyek dapat dicari dengan menggunakan kata tanya *siapa* atau *apa*. Kata tanya *siapa* digunakan untuk mencari subyek yang berupa orang atau sesuatu yang bernyawa. Sedangkan kata tanya *apa* digunakan untuk mencari subyek yang bukan berupa orang atau sesuatu yang tidak bernyawa. Subyek dalam bahasa Indonesia biasanya berupa nomina atau frasa nominal (badanbahasa.kemendikbud.go.id, 2020).

b. Ciri-ciri Predikat

Predikat merupakan salah satu fungsi di dalam kalimat yang merupakan bagian klausa yang menjadi unsur utama di dalam kalimat. Predikat dalam bahasa Indonesia dapat berupa kata kerja (*verba*) atau kelompok kata kerja (*frasa verbal*), kata sifat (*adjektiva*) atau kelompok kata sifat (*frasa adjectival*), atau kata benda (*nomina*) atau keompok kata benda (*frasa nominal*) (badanbahasa.kemendikbud.go.id, 2020).

c. Ciri-ciri Obyek

Objek merupakan salah satu fungsi di dalam kalimat yang kehadirannya bergantung pada jenis predikatnya. Objek biasanya berupa nomina, frasa nominal, atau klausa yang selalu muncul di sebelah kanan predikat yang berupa kata kerja transitif (verba transitif). Jika predikat bukan berupa verba transitif, objek tidak hadir (tidak muncul) di dalam kalimat tersebut (badanbahasa.kemendikbud.go.id, 2020).

d. Ciri-ciri Pelengkap

Pelengkap seperti halnya objek adalah unsur kalimat yang kehadirannya juga bergantung pada predikat. Pelengkap dapat berupa nomina atau frasa nominal, verba atau frasa verbal, dan adjektiva atau frasa adjektiva. Posisi pelengkap dapat terletak di sebelah kanan (setelah atau di belakang) objek atau terletak langsung di sebelah kanan predikat. Jika predikat berupa kata kerja transitif, pelengkap terletak di sebelah kanan objek. Namun, jika predikat bukan berupa kata kerja transitif, mungkin berupa kata kerja intransitif atau berupa kata kerja pasif, pelengkap terletak langsung di sebelah kanan predikat (badanbahasa.kemendikbud.go.id, 2020).

e. Ciri-ciri Keterangan

Keterangan (K) adalah unsur kalimat yang kehadirannya bersifat tidak wajib (opsional). Keterangan dapat berupa nomina (frasa nominal), frasa numeral, berupa frasa preposisional, atau berupa adverbial. Nomina atau frasa nominal yang dapat

menduduki fungsi keterangan biasanya berupa nomina temporal atau nomina yang menyatakan waktu. Selain itu, keterangan dapat dibedakan menjadi dua, yaitu keterangan wajib (wajib hadir/wajib muncul dalam kalimat) dan keterangan manasuka. Keterangan wajib merupakan bagian dari predikat, sedangkan keterangan manasuka bukan bagian dari predikat. Keterangan manasuka merupakan keterangan yang sejajar dengan subjek dan predikat (badanbahasa.kemendikbud.go.id, 2020).