

## BAB 2

### LANDASAN TEORI

#### 2.1 Aturan Penggunaan Huruf Miring

Penggunaan huruf miring dalam tata tulis Bahasa Indonesia bukan sekadar pilihan stilistika, melainkan sebuah perangkat tipografis yang diatur secara formal untuk menjalankan fungsi linguistik yang spesifik. Pedoman Umum Ejaan Bahasa Indonesia (PUEBI) menjadi acuan tertinggi yang membakukan kaidah ini, memastikan adanya konsistensi dan standar dalam komunikasi tulis [13]. PUEBI secara sistematis menguraikan tiga fungsi utama penggunaan huruf miring, yang masing-masing memiliki tujuan komunikatif yang berbeda. Aturan-aturan ini menjadi landasan bagi penulis, editor, dan sistem pemeriksaan ejaan dalam memastikan kepatuhan terhadap standar kebahasaan.

Fungsi pertama dalam penggunaan huruf miring adalah untuk penulisan judul karya. Huruf miring wajib digunakan saat menuliskan judul buku, judul film, judul album lagu, judul acara televisi, judul sinar, judul lakon, dan nama media massa yang dikutip dalam tulisan [13]. Sebagai contoh, dalam kalimat "Berita itu muncul dalam surat kabar *Republika*" atau "Saya sudah membaca buku *Salah Asuhan* karangan Abdoel Moeis" [14]. Penggunaan huruf miring secara jelas memisahkan judul karya dari sisa kalimat, sehingga pembaca dapat dengan mudah mengidentifikasi sebuah entitas rujukan.

Fungsi kedua dalam penggunaan huruf miring adalah untuk penegasan. Huruf miring dapat digunakan untuk memberikan penekanan pada huruf, bagian dari suatu kata, kata atau kelompok kata tertentu dalam sebuah kalimat [13]. Tujuannya adalah untuk menarik perhatian pembaca pada *element* tersebut. Contoh dari fungsi ini adalah "Huruf terakhir kata abad adalah *d*" [13]. Contoh tersebut menunjukkan bahwa huruf d dibuat menjadi huruf miring untuk menandai huruf terakhir dari kata abad.

Fungsi ketiga dalam penggunaan huruf miring adalah untuk penulisan istilah asing atau bahasa daerah. PUEBI menetapkan bahwa kata atau ungkapan yang berasal dari bahasa selain Indonesia, termasuk nama-nama ilmiah, harus ditulis dengan huruf miring [13]. Hal ini bertujuan untuk menandai bahwa kata tersebut belum sepenuhnya terserap ke dalam Kamus Besar Bahasa Indonesia (KBBI). Contohnya yaitu "Nama ilmiah buah manggis adalah *Gracinia mangostana*",

”*Weltanschauung* bermakna ‘pandangan dunia’.” [13].

Namun, terdapat pengecualian penting dalam aturan penggunaan huruf miring. Penggunaan huruf miring untuk istilah asing tidak berlaku untuk nama diri, seperti nama orang, lembaga, organisasi, atau merek dagang dalam bahasa asing atau bahasa daerah tidak ditulis dengan huruf miring [13]. Misalnya, nama Steve Rogers atau lembaga Yamaha Music School tidak ditulis dengan huruf miring, meskipun bukan berasal dari Bahasa Indonesia [15].

## 2.2 *Natural Language Processing*

*Natural Language Processing* (NLP) merupakan cabang dari kecerdasan buatan (AI) yang berfokus pada pengembangan sistem yang mampu memahami dan memproses bahasa manusia secara alami [16]. Bidang ini berada pada irisan antara linguistik, ilmu komputer, dan pembelajaran mesin, yang bersama-sama bertujuan memungkinkan mesin menafsirkan dan menghasilkan bahasa manusia secara bermakna [16].

Dalam tinjauan sistematis, NLP diidentifikasi sebagai salah satu bidang paling populer dalam ranah kecerdasan buatan [17]. Peningkatan popularitas tersebut dipicu oleh pertumbuhan pesat data tekstual digital dan meningkatnya kebutuhan otomatisasi analisis bahasa dalam berbagai domain [16] [18]. Aplikasi NLP kini tidak hanya berfokus pada semantik dan linguistik, tetapi juga telah meluas ke bidang-bidang lain seperti keamanan siber, kimia, studi gender, bibliometrika, dan bahkan mekanika kuantum [17]. Hal ini menunjukkan bahwa NLP telah berkembang melampaui konteks linguistik tradisional menuju penerapan multidisipliner yang luas.

Evolusi NLP dapat ditelusuri dari pendekatan berbasis aturan menuju metode statistik, dan kini mencapai era pembelajaran mendalam dengan model pra-latih seperti BERT dan GPT [16] [18]. Pergeseran ini ditandai dengan munculnya paradigma baru yang mengintegrasikan berbagai kerangka pemodelan seperti *sequence labeling*, *sequence-to-sequence*, dan *language modeling*, yang memungkinkan satu model menangani beragam tugas bahasa alami [19]. Fenomena tersebut dikenal sebagai *paradigm shift* dan menjadi tonggak penting dalam penyatuan berbagai tugas NLP.

Meskipun kemajuan model bahasa sangat pesat, tantangan mendasar masih muncul pada aspek generalisasi. Penelitian menunjukkan bahwa performa tinggi model pada data uji konvensional yang umumnya diasumsikan bersifat *independent*

*and identically distributed* (i.i.d.) belum tentu mencerminkan kemampuan generalisasi terhadap data baru dengan distribusi yang berbeda [20]. Kondisi ini menuntut pengembangan taksonomi dan metode evaluasi baru yang mampu mengukur sejauh mana model NLP dapat mentransfer pengetahuan secara andal di berbagai konteks [20].

Secara keseluruhan, NLP kini telah bertransformasi menjadi disiplin yang matang dan dinamis. Perpaduan antara pembelajaran mendalam, ketersediaan data besar, serta kerangka evaluasi yang lebih ketat menempatkan NLP sebagai salah satu pilar utama perkembangan kecerdasan buatan modern [16] [17] [18] [19] [20]. Salah satu pendekatan yang banyak digunakan dalam berbagai penelitian NLP adalah pelabelan sekuensial (*sequence labeling*). Pendekatan ini digunakan untuk memberikan label atau kategori pada setiap satuan teks, seperti kata atau frasa, dan menjadi dasar bagi beragam tugas penting dalam NLP, termasuk penandaan kelas kata, pengenalan entitas bernama, serta ekstraksi informasi [19].

### **2.2.1 Pelabelan Sekuensial (*Sequence Labeling*) sebagai Tugas Fundamental**

Banyak permasalahan inti dalam NLP, termasuk tugas deteksi kesalahan penggunaan huruf miring dalam penelitian ini, dapat diformulasikan sebagai tugas pelabelan sekuensial (*sequence labeling*). Tugas ini melibatkan proses pemberian label kategoris pada setiap unit atau token, seperti kata atau karakter, dalam suatu urutan teks, misalnya kalimat [21] [22]. Pelabelan sekuensial merupakan salah satu pendekatan mendasar dalam ekstraksi informasi yang berfungsi untuk mengubah teks tidak terstruktur menjadi data terstruktur yang dapat diolah lebih lanjut.

Dalam penerapannya, pelabelan sekuensial digunakan secara luas dalam berbagai jenis tugas NLP, termasuk *Part-of-Speech (POS) Tagging* dan NER [22]. *POS Tagging* merupakan proses memberikan penanda kategori gramatikal pada setiap kata dalam suatu teks dan dianggap sebagai komponen penting dalam berbagai aplikasi NLP, seperti analisis sintaksis, ekstraksi informasi, dan penerjemahan mesin [23]. Sementara itu, NER bertujuan untuk mengidentifikasi serta mengklasifikasikan entitas bernama dalam teks, seperti nama orang (*PERSON*), organisasi (*ORGANIZATION*), dan lokasi (*LOCATION*) [17].

Selain itu, pelabelan sekuensial juga digunakan dalam tugas *Slot Filling* pada sistem dialog berorientasi tugas (*task-oriented dialogue systems*). Tugas ini berfungsi untuk mengekstraksi informasi penting dari ucapan pengguna dengan cara memberikan label pada setiap bagian kalimat sesuai kategori yang relevan,

misalnya mengenali “Jakarta” sebagai tujuan perjalanan (*DESTINATION*) dan “besok” sebagai waktu keberangkatan (*DATE*) [24].

### 2.3 IOB (*Inside, Outside, Beginning*) Labeling

IOB (*Inside, Outside, Beginning*) atau yang biasa disebut BIO merupakan metode pelabelan yang digunakan dalam tugas pelabelan sekuensial untuk menandai posisi setiap token dalam entitas. Metode ini memungkinkan model mempelajari batas awal, bagian dalam, dan luar suatu entitas pada teks [25].

Skema IOB terdiri dari tiga label utama. Label B-ENTITAS digunakan untuk menandai token yang berada di awal suatu entitas. Label I-ENTITAS digunakan untuk menandai token yang merupakan bagian dalam dari entitas yang sama, sedangkan label O digunakan untuk token yang berada di luar entitas [25] [26]. Skema IOB menetapkan bahwa tag I-ENTITAS hanya dapat mengikuti B-ENTITAS atau I-ENTITAS dengan tipe yang sama [26]. Ketentuan ini digunakan untuk menjaga konsistensi struktur label dan menghindari urutan yang tidak valid, seperti transisi O yang diikuti oleh I-ENTITAS.

Pendekatan pelabelan ini digunakan secara luas dalam berbagai bidang pemrosesan bahasa alami. Misalnya, penelitian [25] menunjukkan bahwa format IOB dapat diterapkan pada tingkat karakter untuk meningkatkan akurasi identifikasi batas entitas pada teks berbahasa Mandarin. Dalam bidang pengenalan entitas bernama, skema ini juga digunakan dalam berbagai model berbasis BiLSTM dan CRF [22] [27] [28].

Sebagai contoh, kalimat “Saya sudah membaca buku Salah Asuhan.” dapat dianotasi menggunakan skema IOB menjadi: Saya (O) sudah (O) membaca (O) buku (O) Salah (B-JUDUL) Asuhan (I-JUDUL). (O)

### 2.4 *Conditional Random Field* (CRF)

*Conditional Random Field* (CRF) merupakan model grafis probabilistik yang banyak digunakan untuk tugas prediksi terstruktur, terutama pelabelan sekuensial seperti *Part-of-Speech (POS) Tagging* dan *Named Entity Recognition* (NER) [29] [30]. Model ini dikembangkan oleh Lafferty, McCallum, dan Pereira (2001) sebagai pendekatan diskriminatif yang secara langsung memodelkan probabilitas kondisional urutan label  $Y = (y_1, \dots, y_n)$  terhadap urutan observasi  $X = (x_1, \dots, x_n)$ , yaitu  $P(Y|X)$ , tanpa harus mengasumsikan independensi fitur seperti pada model

generatif (*Hidden Markov Model*) [29] [31]. Pendekatan ini memungkinkan CRF memanfaatkan fitur-fitur kompleks dan tumpang tindih yang relevan dengan konteks data.

Untuk tugas pelabelan sekuensial, jenis CRF yang paling umum digunakan adalah *Linear-Chain CRF*, yang didefinisikan sebagai:

$$P(Y|X) = \frac{1}{Z(X)} \exp \left( \sum_{i=1}^n \sum_k \lambda_k f_k(y_{i-1}, y_i, X, i) \right) \quad (2.1)$$

dengan  $f_k$  merupakan fungsi fitur yang menggambarkan hubungan antara token input dan labelnya,  $\lambda_k$  bobot fitur yang dipelajari selama pelatihan, dan  $Z(X)$  adalah faktor normalisasi yang menjamin total probabilitas bernilai satu [29] [31]. Bobot positif menunjukkan dukungan terhadap transisi antar label tertentu, sedangkan bobot negatif menunjukkan sebaliknya.

Persamaan (2.1) menggambarkan fungsi probabilitas kondisional dari model *Conditional Random Field* (CRF) yang bergantung pada parameter bobot  $\lambda_k$ . Selama proses pelatihan, parameter-parameter dalam model CRF dioptimalkan menggunakan algoritma *Limited-memory Broyden-Fletcher-Goldfarb-Shanno* (L-BFGS), yang efisien untuk data berskala besar tanpa menyimpannya secara penuh [32].

Selama pelatihan, parameter  $\lambda_k$  disesuaikan untuk memaksimalkan probabilitas *log-likelihood* dari data anotasi, sebagaimana dirumuskan dalam Persamaan 2.2:

$$\log P(Y|X) = \sum_{i=1}^n \sum_k \lambda_k f_k(y_{i-1}, y_i, X, i) - \log Z(X) \quad (2.2)$$

Proses ini umumnya dilakukan menggunakan algoritma optimisasi seperti *Limited-memory BFGS* (L-BFGS), yang efisien untuk dataset berskala besar [33].

Pada tahap inferensi, CRF tidak menghitung semua kombinasi label secara eksplisit karena kompleksitasnya bersifat eksponensial. Sebagai gantinya, digunakan algoritma *Viterbi*, yaitu algoritma pemrograman dinamis yang secara efisien menemukan urutan label dengan skor global tertinggi [31]. Penggunaan algoritma ini memastikan hasil prediksi tidak hanya optimal secara lokal di setiap token, tetapi juga konsisten secara keseluruhan terhadap konteks sekuens.

Dalam arsitektur NLP modern, CRF sering dipadukan dengan jaringan saraf seperti *Bidirectional Long Short-Term Memory* (BiLSTM) atau model berbasis



*Attention*, membentuk arsitektur *BiLSTM-CRF* [34]. Lapisan *BiLSTM* bertugas mengekstraksi representasi kontekstual dari urutan kata, sedangkan lapisan *CRF* berfungsi menjaga konsistensi struktur label, misalnya dengan menjamin agar label *Inside* hanya muncul setelah label *Beginning* [31] [34]. Kombinasi ini terbukti memberikan kinerja unggul pada berbagai tugas pelabelan sekuensial, termasuk pengenalan entitas hukum dan teks biomedis.

## 2.5 Confusion Matrix

Evaluasi performa model klasifikasi merupakan tahap krusial dalam pengembangan sistem pembelajaran mesin karena menentukan kemampuan model dalam mengenali pola dan melakukan generalisasi terhadap data baru. Salah satu metode evaluasi yang paling umum digunakan adalah *confusion matrix*, yaitu tabel yang menggambarkan hubungan antara label aktual (*ground truth*) dan label prediksi model [35] [36]. Untuk kasus klasifikasi biner, *confusion matrix* berbentuk  $2 \times 2$  dan terdiri atas empat komponen utama: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [35] [37]. Keempat komponen ini menjadi dasar bagi perhitungan metrik evaluasi kuantitatif yang digunakan untuk menilai kinerja keseluruhan model.

*True Positive* (TP) didefinisikan sebagai jumlah kasus ketika model berhasil memprediksi kelas positif secara benar, sedangkan *True Negative* (TN) menunjukkan jumlah kasus ketika model secara tepat mengidentifikasi kelas negatif [35] [37]. Kedua komponen ini merepresentasikan prediksi yang sesuai antara label aktual dan hasil klasifikasi model.

Sebaliknya, *False Positive* (FP) menggambarkan kondisi ketika model secara keliru memprediksi data negatif sebagai positif, yang dikenal sebagai kesalahan tipe I. Sementara itu, *False Negative* (FN) terjadi ketika model gagal mengenali kelas positif dan secara salah mengklasifikasikannya sebagai negatif, atau disebut kesalahan tipe II [35] [37]. Keempat komponen ini menjadi dasar utama dalam evaluasi performa model klasifikasi, karena memungkinkan analisis terhadap pola kesalahan serta konsistensi prediksi yang dihasilkan oleh model [36] [37].

*Confusion matrix* merupakan dasar perhitungan sejumlah metrik evaluasi kuantitatif, seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, yang secara komplementer menggambarkan performa keseluruhan model klasifikasi [37]. Penggunaan matriks ini memungkinkan pemetaan hasil prediksi ke dalam format visual yang memudahkan analisis distribusi kesalahan secara lebih intuitif

dibandingkan sekadar angka ringkasan. Contoh tabel sederhana *confusion matrix* dapat dilihat pada Table 2.1.

Tabel 2.1. Tabel *Confusion Matrix*

	<i>Actual Positive</i>	<i>Actual Negative</i>
<i>Predicted Positive</i>	TP	FP
<i>Predicted Negative</i>	FN	TN

Metrik *Accuracy* mengukur proporsi total prediksi yang benar terhadap seluruh prediksi yang dilakukan model, sebagaimana dirumuskan pada Persamaan 2.3.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

Meskipun *accuracy* sering digunakan karena sifatnya yang intuitif, metrik ini dapat menjadi kurang representatif pada data dengan distribusi kelas yang tidak seimbang [35] [37].

Untuk memberikan evaluasi yang lebih spesifik, digunakan metrik *Precision* dan *Recall*. *Precision* menunjukkan proporsi prediksi positif yang benar-benar termasuk dalam kelas positif, sebagaimana ditunjukkan dalam Persamaan 2.4.

$$Precision = \frac{TP}{TP + FP} \quad (2.4)$$

Sementara itu, *Recall* atau sensitivitas mengukur sejauh mana model berhasil mendeteksi seluruh kasus positif aktual, sebagaimana dirumuskan pada Persamaan 2.5 [37].

$$Recall = \frac{TP}{TP + FN} \quad (2.5)$$

Kedua metrik ini bersifat saling melengkapi *Precision* menekankan minimisasi kesalahan tipe I (*false positives*), sedangkan *Recall* menekankan pengurangan kesalahan tipe II (*false negatives*) [37].

Untuk menyeimbangkan keduanya, digunakan metrik *F1-Score*, yang merupakan rata-rata harmonik antara *Precision* dan *Recall*, sebagaimana ditunjukkan pada Persamaan 2.6.

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.6)$$

*F1-Score* dianggap efektif untuk mengukur performa model pada dataset dengan distribusi kelas yang tidak seimbang karena mempertimbangkan kedua jenis

kesalahan secara simultan [35] [36] [37].



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA