

BAB 2 LANDASAN TEORI

Pada bagian ini dijabarkan beberapa teori yang mendasari penelitian secara lengkap. Penyajian yang diberikan berupa definisi teori-teori yang langsung berkaitan dengan topik yang sedang diteliti, yang dijabarkan sebagai berikut:

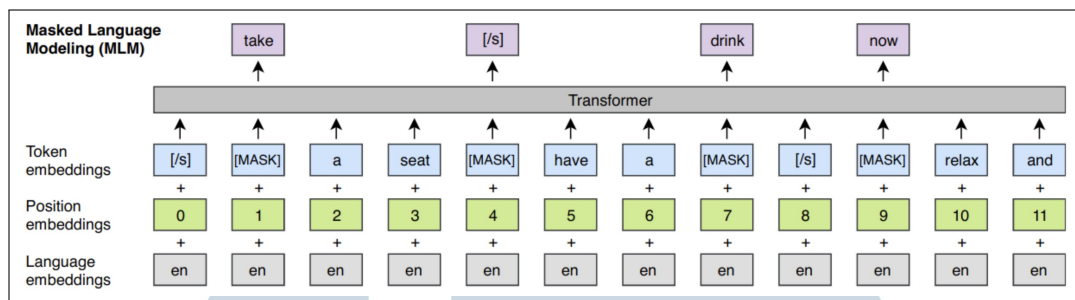
2.1 *Multilingual Text Classification*

Multilingual text classification merupakan salah satu fokus penelitian dalam bidang *Natural Language Processing* (NLP). Berbeda dengan klasifikasi teks *monolingual*, terdapat dua kesulitan dalam melakukan klasifikasi *multilingual*: (1) Sulitnya membagi dan memaknai kata-kata dengan bahasa yang berbeda. (2) Kebanyakan dari model yang ada sekarang hanya fokus pada klasifikasi teks satu bahasa sehingga memiliki tingkat adaptasi yang rendah terhadap teks *multilingual* [18].

Namun, saat ini sudah mulai berkembang penelitian yang memungkinkan untuk melakukan klasifikasi teks *multilingual*. Kebanyakan dari penelitian yang telah dibuat, model dilatih menggunakan dataset yang *monolingual*, kemudian dilakukan *translation mechanism* untuk melakukan klasifikasi teks dengan bahasa yang berbeda [13].

2.2 XLM-R

XLM-R merupakan sebuah model multibahasa *Mask Language Model* berbasis *transformer*. XLM-R sebelumnya sudah dilatih menggunakan teks dalam 100 bahasa, dan memberikan performa yang sangat baik dalam klasifikasi multibahasa, memberikan label, dan juga menjawab pertanyaan. Model ini dikembangkan dengan tujuan untuk membantu riset *Natural Language Processing* (NLP) di bidang multibahasa dengan pemahaman bahasa yang sumber dayanya rendah [9].



Gambar 2.1. Gambaran Masked Language Model

Pengembangan XLM-R difokuskan pada *unsupervised learning*, sehingga pengembangan dilakukan dengan menggunakan metode *Mask Language Model* (MLM). MLM bekerja dengan membuat sebuah *mask* secara acak dari *token input*, yang kemudian melakukan prediksi dari *masked token* tersebut. Teknik MLM ini membantu model dalam mempelajari keterkaitan antara kata dan kalimat. Penggunaan metode MLM pada proses *pre-train* XLM-R membantu model untuk memahami banyak bahasa dan mempelajari *pattern cross-lingual* dari bahasa-bahasa tersebut. Untuk menguji kemampuan model XLM-R dalam pemahaman *cross-lingual*, digunakan evaluasi *benchmark* seperti *cross-lingual natural language inference*, *named entity recognition*, dan *question answering*. XLM-R berhasil mendapatkan hasil yang lebih tinggi dari kebanyakan model sebelumnya untuk tugas-tugas tersebut [9].

2.3 Transfer Learning

Konsep *transfer learning* merupakan sebuah konsep di mana sebuah model *pre-trained* yang sudah memiliki pengetahuan didalamnya, di latih dan diberikan tugas dan data baru yang berbeda. Sehingga model dapat beradaptasi dalam melakukan tugas baru tersebut dengan pengetahuan yang ada [19]. *Transfer learning* memanfaatkan *pre-trained* model yang sebelumnya sudah dilatih pada corpus yang besar. Penggunaan metode *transfer learning* pada *pre-trained* model memiliki nilai *F1-score* yang lebih tinggi dibandingkan dengan model pembelajaran mesin lainnya [20].

2.4 Metrik Evaluasi

Metrik evaluasi merupakan sebuah metode untuk melakukan evaluasi hasil performa dari model yang telah dibuat. Terdapat beberapa metrik evaluasi yang

dapat digunakan untuk mengukur performa sebuah model, di antaranya adalah *accuracy*, *precision*, *recall*, dan *F1-score* [21].

1. *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Akurasi merupakan salah satu metrik yang paling umum digunakan untuk melakukan evaluasi klasifikasi. Untuk melakukan perhitungan *accuracy*, perlu dilakukan penjumlahan dari elemen *True Positive* dengan *True Negative* dan kemudian dibagi dengan semua elemen metrik. Elemen *True Positive* dan *True Negative* merupakan hasil benar yang diprediksi oleh model dan merupakan bagian diagonal utama pada metrik evaluasi.

2. *Precision*

$$Precision = \frac{TP}{TP+FP}$$

Presisi merupakan nilai kebenaran dari prediksi yang dilakukan. *True Positive* merupakan elemen yang telah diberikan label positif oleh model dan hasilnya benar positif, sementara *False Positive* merupakan elemen yang diberikan label positif namun, hasil yang benar adalah negatif. Hasil dari Presisi menyatakan seberapa besar model dapat dipercaya ketika melakukan prediksi dan memberikan hasil yang positif.

3. *Recall*

$$Recall = \frac{TP}{TP+FN}$$

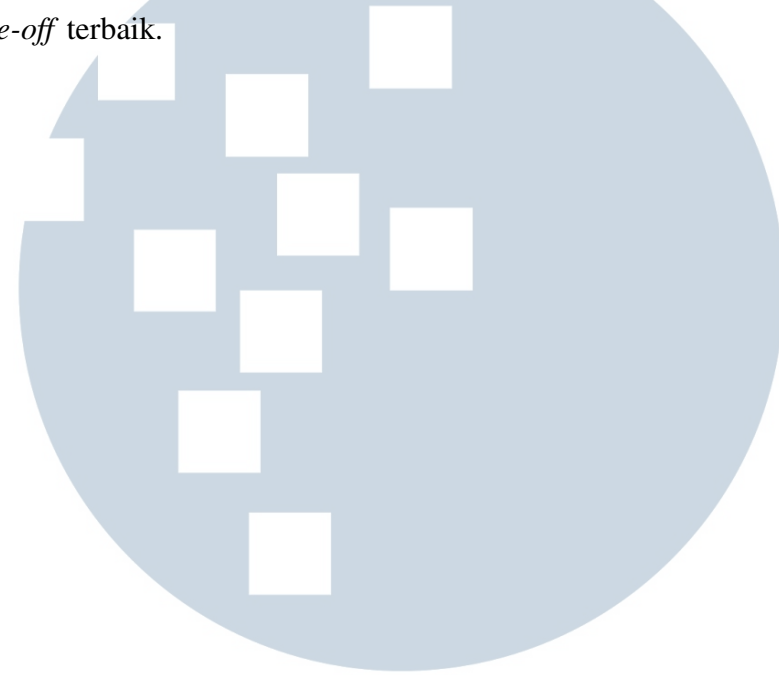
Recall merupakan hasil pembagian dari elemen *True Positive* dengan total dari elemen yang diklasifikasikan sebagai positif yaitu *True Positive* dan *False Negative*. *False Negative* merupakan elemen yang diberikan label negatif, namun sebenarnya positif. *Recall* digunakan untuk mengukur akurasi dari hasil prediksi model untuk kelas positif dan mengukur kemampuan model dalam menemukan semua kelas positif pada kumpulan data.

4. *F1-Score*

$$F1Score = \left(\frac{2}{\frac{1}{precision} + \frac{1}{recall}} \right) = 2 * \left(\frac{precision * recall}{precision + recall} \right)$$

F1-Score merupakan metode yang dapat mengevaluasi performa model klasifikasi dengan melihat nilai dari metrik evaluasi. Metode ini bekerja dengan menggabungkan nilai *precision* dan *recall* dengan menggunakan

konsep rata-rata *harmonic*. Artinya, *F1-Score* menentukan nilai keseimbangan antara kedua nilai tersebut dengan nilai terbaik adalah 1 dan nilai terburuk adalah 0. Kontribusi dari *precision* dan *recall* pada *F1-Score* sama dengan rata-rata *harmonic* yang berguna untuk mendapatkan titik *trade-off* terbaik.



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA