

**DETEKSI DINI PENDERITA DEPRESI DALAM BAHASA INDONESIA
DENGAN PENDEKATAN *FINE-TUNING* MODEL BAHASA DAN FITUR
METADATA LINGUISTIK**

Bab I Pendahuluan

I.1 Latar Belakang

Depresi adalah masalah yang umum terjadi pada manusia dimana sekitar 450 juta orang di dunia pernah mengalami depresi (WHO, 2017). Depresi ditandai dengan adanya perasaan sedih, tertekan, kehilangan minat, perasaan rendah diri gangguan tidur dan kurangnya konsentrasi menyebabkan menurunnya kemampuan berpikir dan mengatasi masalah pada banyak orang (WHO, 2012).

Di Indonesia, gangguan depresi menempati peringkat pertama untuk *mental disorder* yang menyebabkan DALYs (*Disability Adjusted Life Year*), dimana persentase kontributor kematian terbesar disebabkan oleh gangguan mental yakni sekitar 14,4% (Kemenkes RI 2020). Namun, lebih dari 78% penderita gangguan mental di negara dengan penghasilan rendah-menengah termasuk salah satunya adalah Indonesia, masih belum mendapatkan perawatan (Kohn, et al., 2004).

Faktanya, hingga saat ini depresi di deteksi berdasarkan *self-reported* oleh penderita, *behavior report* oleh teman penderita, serta *mental health examination*, yang mana, belum ada laboratorium uji yang dapat mendiagnosa berbagai bentuk gangguan mental. Dalam kehidupan sehari - hari, sosial media memfasilitasi pra-diagnosa kondisi kesehatan mental klinis terkait kecemasan dan gangguan depresi oleh pengguna, yang mengungkapkan dan berbagi kegelisahan (Kumar et al., 2019); (Joshi & Patwardhan, 2020)

Penelitian untuk mendeteksi depresi dari *tweet* untuk Bahasa Indonesia sudah pernah dilakukan, yakni dengan pendekatan leksikal oleh Oyong et al.(2018). Oyong et al. (2018) menggunakan istilah-istilah gejala depresi dan kemunculan

leksikal depresi untuk melakukan pembobotan dan menghasilkan skor depresi dari teks yang diselesaikan dengan Pengolahan Bahasa Alami (Oyong et al., 2018). Eksperimen yang dilakukan pada penelitian ini adalah untuk membandingkan akurasi dari *preprocessing* standar, dan *preprocessing* yang khusus untuk teks dari sosial media, yakni Twitter. Akurasi dari penelitian ini yakni, untuk nilai F1 dari *preprocessing* standar dan *text-specific preprocessing* memiliki nilai yang sama yakni 0.47, dan nilai sensitivitas yang sama pula yakni 0.89 dengan nilai ambang 0.5. Sedangkan nilai F1 yang lebih baik yakni 0.5 untuk *text-specific preprocessing* pada nilai ambang 0.8. Nilai threshold ini ditentukan dari hasil penghitungan nilai depresi yang merupakan jumlah dari perkalian bobot *pronoun*, bobot negasi, frekuensi dan gejala. Berdasarkan hasil tersebut klasifikasi penderita depresi masih belum baik, hal ini diduga karena pendekatan leksikal yang digunakan memungkinkan untuk terjadi masalah OOV, dimana tingkat token yang mendalam dan pemahaman terhadap sebuah kalimat sangat dibutuhkan apalagi untuk domain spesifik (Wang et al., 2019). Selain itu, penggunaan metode penghitungan kemunculan leksikal mengharuskan untuk mendaftarkan semua daftar kata gejala depresi yang ada sehingga akan menjadi tidak efisien sehingga dibutuhkan sebuah model yang mampu melakukan generalisasi terhadap gejala depresi sehingga dapat meningkatkan performa model yang akan dibangun.

Untuk mengatasi kesalahan dalam melakukan klasifikasi yang disebabkan oleh OOV, dapat ditangani dengan menerapkan pendekatan pembelajaran transfer (*transfer learning*) (Nayak et al., 2020; Tai et al., 2019) dari suatu model bahasa *pre-trained*. Pembelajaran transfer yang dilakukan akan melakukan *pre-training* untuk bahasa Inggris dari data CLEF eRisks 2017 dan melakukan prediksi untuk dataset berbahasa Indonesia. Berdasarkan penelitian dalam beberapa tahun terakhir, penelitian yang menggunakan pendekatan *transfer learning* pada beberapa *task* seperti *question answering*, *sentiment analysis*, serta *semantic textual similarity* menunjukkan kinerja *state-of-the-art* (Azhar, 2020). Hingga saat ini *pre-train* model

bahasa untuk *cross-lingual* memiliki beberapa varian seperti XLM-RoBERTa dan Model BERT Multi-bahasa (Devlin et al., 2019).

Untuk meningkatkan kinerja model yang dibangun, Trotzek et al., (2020) dan Sisk, (2018) dalam melakukan deteksi dini penderita depresi dengan melakukan klasifikasi berbasis teks dengan metode *pembelajaran mesin*. Arsitektur yang diusulkan oleh Trotzek et al., (2020) telah mencapai kinerja *state-of-the-art* yang mana menggunakan pendekatan berbasis *deep learning*, yakni *Convolutional Neural Network (CNN)* yang dikombinasikan dengan *word embedding* dan metadata linguistik pada level pengguna.

Penelitian tesis ini bertujuan untuk melakukan deteksi depresi dengan mengadopsi arsitektur yang diusulkan oleh Trotzek et al., (2020) dengan menggunakan metadata linguistik level pengguna dengan memanfaatkan pembelajaran transfer dan pendekatan deep learning. Model depresi deteksi pada penelitian yang dilakukan oleh (Oyong et al., 2018) akan dijadikan *baseline* pada penelitian ini sebagai pembanding untuk menguji performa dari pendekatan yang digunakan pada penelitian ini.

I.2 Masalah Penelitian

Masalah dari penelitian ini adalah bagaimana mendeteksi depresi dengan mengadaptasi cara penyelesaian masalah dengan *transfer learning* dengan memanfaatkan linguistik metadata pada level pengguna, yang diusulkan pada penelitian Trotzek et al., (2020) untuk melakukan penanganan OOV dan generalisasi model pada penelitian sebelumnya.

I.3 Tujuan

Tujuan dari penelitian ini adalah untuk melakukan deteksi dini gejala depresi dengan *transfer learning* dengan memanfaatkan metadata linguistik pada level pengguna

I.4 Hipotesis

Premis-1 : Pendekatan berbasis leksikon dapat digunakan untuk melakukan deteksi penderita depresi melalui data Sosial Media (Oyong et al., 2018)

Premis-2 : Pendekatan berbasis leksikon dapat digunakan untuk melakukan deteksi penderita depresi melalui data Sosial Media (Oyong et al., 2018)

Berdasarkan premis-1 disusun hipotesis :

Hipotesis-1: Dengan memanfaatkan metode model bahasa *pre-train kontekstual* akan meningkatkan efektifitas (El-Alami et al., 2021; Wilie et al., 2020) dari berbagai *task* Pemrosesan Bahasa Alami (NLP), baik untuk *high-resource* maupun *low-resource* (Ikhwantri, 2019).

Berdasarkan premis-2 disusun hipotesis :

Hipotesis-2: Agar model dapat melakukan klasifikasi dari gejala depresi pada teks tertulis dengan tepat, maka dapat digunakan metadata linguistik pada level pengguna yang mana telah mencapai *state-of-the-art*.

I.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

- 1) Teks yang digunakan adalah teks berbahasa Inggris dari CLEF eRisks 2017 dan berbahasa Indonesia dengan kata yang mengandung indikasi depresi dari penelitian Oyong et al. (2018)
- 2) Sedangkan model bahasa *pre-train* yang digunakan adalah XLM-RoBERTa dan Bert Multilingual.

I.6 Metodologi

Metodologi penggerjaan penelitian tesis ini adalah sebagai berikut.

1) Analisis Persoalan dan solusi

Pada tahap analisis persoalan yang dihadapi Oyong et al. (2018) dalam melakukan deteksi depresi, kemudian dilakukan studi literatur alternatif solusi dari masalah tersebut.

2) Konstruksi Korpus

Pada tahap ini akan dilakukan konstruksi korpus yang akan digunakan untuk mendeteksi depresi. Korpus yang digunakan adalah data berbahasa inggris dari CLEF 2017 dan data yang diambil dari twitter, yang di cari dengan kata yang terkait dengan depresi pada penelitian yang dilakukan oleh Oyong et al. (2018).

3) Desain Eksperimen

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

4) Implementasi

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

5) Eksperimen

Pada tahap eksperimen akan dilakukan eksperimen untuk mengujikan efektivitas penggunaan model deteksi depresi dengan fitur-fitur yang sudah didefinisikan. Kemudian, akan dilakukan analisis terkait hasil eksperimen.

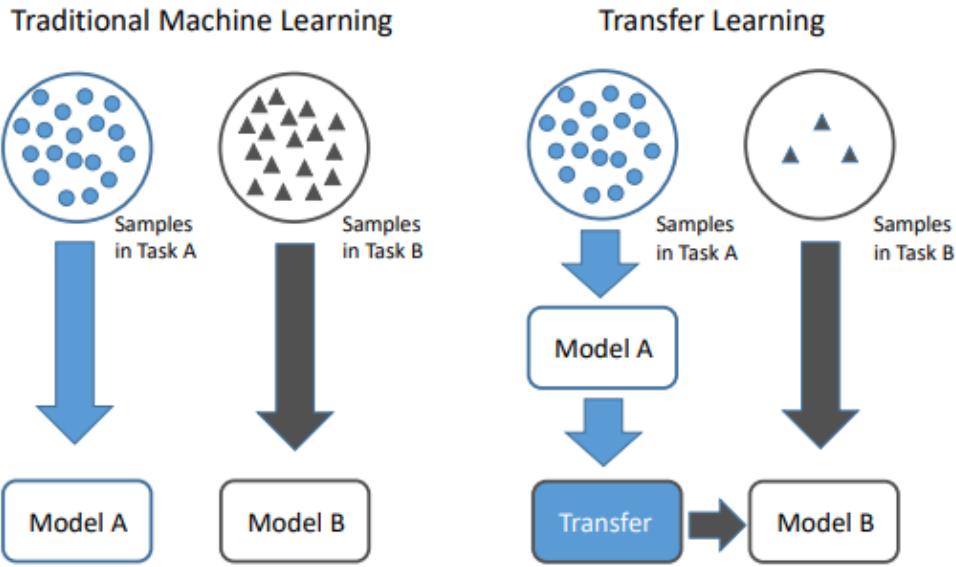
Bab II Tinjauan Pustaka

II. 1 Pembelajaran Transfer di Pemrosesan Bahasa Alami

Saat ini algoritma pembelajaran mesin yang digunakan untuk menyelesaikan *task* di bidang Pemrosesan Bahasa Alami telah menunjukkan kinerja yang sangat baik terutama dalam penggunaan *deep neural network* seperti RNN (*Recurrent Neural Network*), LSTM (*Long Short Term Memory*), GRU (*Gated Recurrent Unit*), maupun CNN (*Convolutional Neural Network*). Algoritma tersebut dapat bekerja dengan sangat baik jika tersedia data beranotasi yang cukup banyak atau yang dikenal dengan Pembelajaran Terbimbing (*Supervised Learning*) yang mana model ini sensitif terhadap *noise* dan rentan mengalami *overfitting* (Ruder, 2019).

Pembelajaran transfer dapat menjadi salah satu solusi dari permasalahan diatas dimana pembelajaran transfer akan memanfaatkan pengetahuan dari domain sumber dan menerapkannya pada *tasks* target dan domain target (Pan and Yang, 2010; Pan, 2014; Ruder, 2019). Secara formal, Pan and Yang (2010) menjelaskan tentang definisi dari pembelajaran transfer yakni untuk domain sumber D_s , dengan *tasks* sumber T_s , kemudian domain target D_t , dan *tasks* target T_t , objektif dari pembelajaran transfer adalah untuk mempelajari *conditional probability distribution* dari $P_t(Y_t | X_t)$ dengan X adalah ruang fitur dan Y adalah himpunan label. Probabilitas tadi dihitung berdasarkan informasi yang didapat dari D_s dan D_t dimana $D_s \neq D_t$ atau $T_s \neq T_t$, dengan jumlah data teranotasi yang terbatas ataupun dengan data tidak teranotasi dengan jumlah yang tidak terbatas.

Ilustrasi perbandingan paradigm dari pembelajaran terbimbing tradisional dan pembelajaran transfer dapat dilihat pada gambar II.1.



Gambar II.1. Ilustrasi Proses Pembelajaran Mesin Tradisional dan Pembelajaran Transfer (Ruder, 2019)

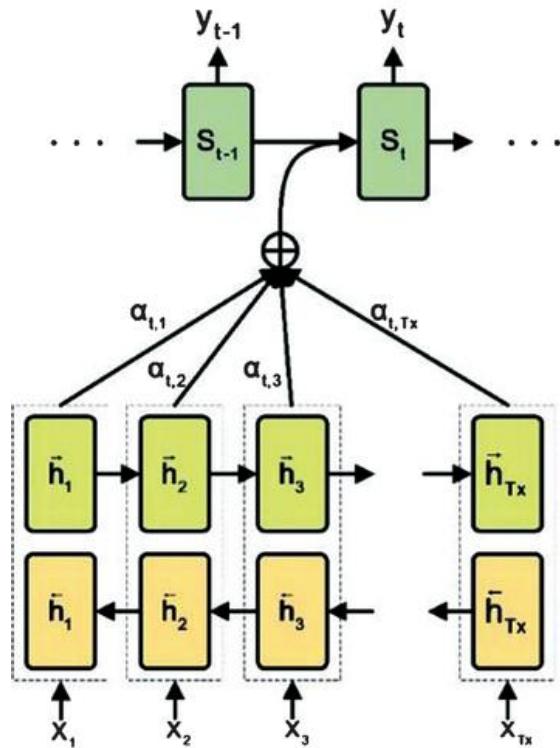
Kemudian, untuk task yang menggunakan *Pre-train Contextual Language Model* sudah banyak digunakan dalam beberapa *task NLP* (Peters et al., 2018; Devlin et al., 2019) dan menunjukkan hasil yang mencapai *state-of-the-art* pada berbagai *tasks* di NLP (Conneau et al., 2019; Yang et al., 2017; El-Alami et al., 2021), termasuk untuk Bahasa Indonesia yang merupakan bahasa *low-resource* (Wilie et al., 2020).

II. 2 Mekanisme Atensi dan Transformer

Pendekatan konvensional dari *neural machine translation*, disebut dengan encoder-decoder, yang mengkodekan kalimat masukan kedalam *fixed-length vector*. Bahdanau et al., (2015) memperkirakan pendekatan konvensional ini akan menimbulkan masalah ketika menerjemahkan kalimat yang panjang karena harus mengompres kalimat *input* kedalam *fixed-length vector* sehingga, *neural network* akan mengabaikan informasi penting yang ada dalam kalimat input (Cho et al., 2015).

Arsitektur dari *neural network* yang biasa digunakan untuk menyelesaikan kasus *machine translation* adalah LSTM (*Long-short Term Memory*) arsitektur ini adalah integrasi dari RNN (*Recurrent Neural Network*) yang mampu mengatasi masalah *long-range dependence* dari RNN. Namun, LSTM dimungkinkan masih belum mampu bekerja dengan baik khususnya jika kalimat masukan lebih panjang dari pada kalimat pada korpus latihnya.

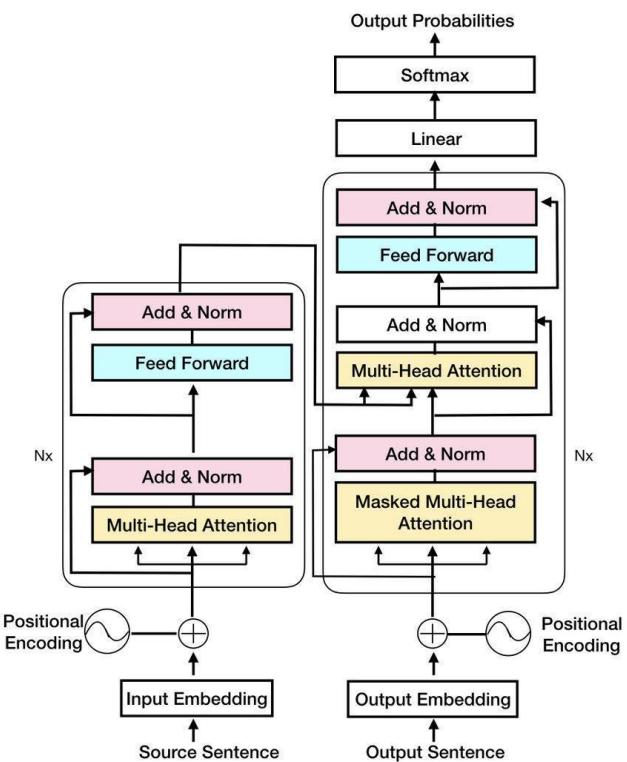
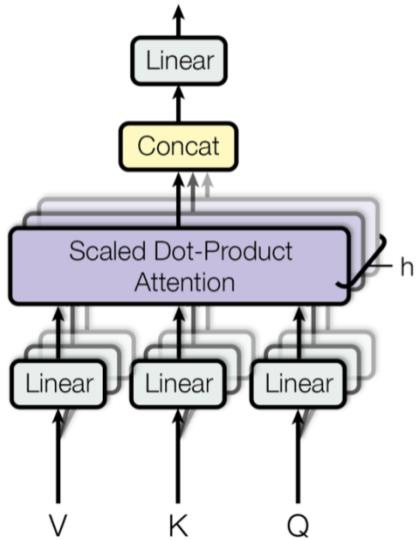
Untuk mengatasi masalah tersebut, Bahdanau et al., (2015) mengusulkan mekanisme *Attention* yang memiliki kemampuan untuk fokus pada elemen yang spesifik pada data. Perbedaan utama dari pendekatan ini dengan *encoder-decoder* konvensional ialah pada mekanisme atensi ini tidak dilakukan *encoding* ke dalam *fixed-length vector* melainkan melakukan *encoding* terhadap kalimat masukan ke dalam vektor sekuensial dan memilih subset dari vektor tersebut selama proses translasi *decoding* (Bahdanau et al., 2015). Ilustrasi dari mekanisme atensi ini data dilihat pada gambar II.2.



Gambar II.2 Ilustrasi Grafis dari Mekanisme *Attention* (Bahdanau et al., 2015)

Vektor konteks C_i mengacu pada label sekuensial $(h_1, h_2, h_3, \dots, h_{Tx})$, dimana masing - masing label h_i mengandung informasi tentang keseluruhan masukan dengan berfokus pada bagian – bagian pada kata ke - i dari kalimat masukan.

Mekanisme atensi telah terbukti sebagai integrasi dari pemodelan sekuensial dan pemodelan transduksi yang memungkinkan pemodelan tanpa mempertimbangkan jarak symbol pada urutan input dan output. Namun, dalam beberapa kasus, mekanisme ini dilakukan berulang. Sehingga, untuk mengatasi masalah itu Vaswani et al., (2017) mengusulkan *Transformer* yang mampu melihat keseluruhan input sekaligus. Pada *Transformer* modul *Attention* mengulang komputasi beberapa kali secara paralel dan disebut dengan *Head Attention*. Kemudian modul *Attention* membagi parameter *Query*, *Key*, dan *Value* dan melewati masing – masing parameter yang dibagi tadi secara independen pada *Head* yang terpisah sehingga disebut *Multi-head Attention*. Ilustrasi dari *Multi-head Attention* dan *Transformer* dapat dilihat pada gambar II.3.



Gambar II.3. *Multi-head Attention* (kiri) dan Arsitektur *Transformer* (kanan)

II. 3 BERT (Bidirectional Encoder from Transformer)

Saat ini perkembangan empiris terhadap *tasks* NLP salah satunya disebabkan adanya teknik *Transfer Learning*, dimana model *deep learning* dilatih dengan data yang sangat besar kemudian digunakan untuk menyelesaikan *tasks* yang mirip dengan data yang berbeda. Salah satu model representasi bahasa yang menerapkan pendekatan ini adalah BERT yang memanfaatkan konsep *encoder* pada transformer.

BERT didesain untuk membaca sekuens input secara dua arah (*bidirectional*), berbeda dengan teknik pemodelan bahasa sebelumnya yang membaca input sekuens hanya dalam satu arah (*unidirectional*) misalnya, pada OpenAI GPT yang membaca token dari kiri ke kanan sehingga token hanya dapat membaca token sebelumnya. Hal ini membuat BERT menjadi lebih baik dalam memahami konteks kalimat, dan mencapai *state-of-the-art* seperti pada *Question Answering and Language Inference* (Devlin et al., 2019) hanya dengan mengganti layer output.

Saat ini, pengembangan BERT telah menghasilkan beberapa varian yang dapat dibedakan misalnya berdasarkan ukuran modelnya, domain aplikasi, dan bahasa korpus yang digunakan selama pelatihan. Saat ini BERT memiliki 2 ukuran yakni *base* dengan 12 *layers*, 768 *hidden size*, 12 *self-attention heads*, dan 110 juta parameter, dan *large* yang memiliki 24 *layers*, 1024 *hidden size*, 16 *self-attention heads*, dan 340 juta parameter. Selain itu, beberapa varian lain seperti, SciBert (*Scientific*), BioBert (*Biomedical text*), dan ClinicalBERT (*clinical notes*), Multilingual BERT, XLM-RoBERTa dan IndoBERT (Wilie et al., 2020).

Kerangka kerja pada BERT terdiri dari 2 langkah, yakni *pre-training* dan *fine-tuning*. Selama melakukan *pre-training*, BERT melakukan 2 *unsupervised tasks* yakni sebagai berikut :

1. *Masked Language Modelling* (MLM)

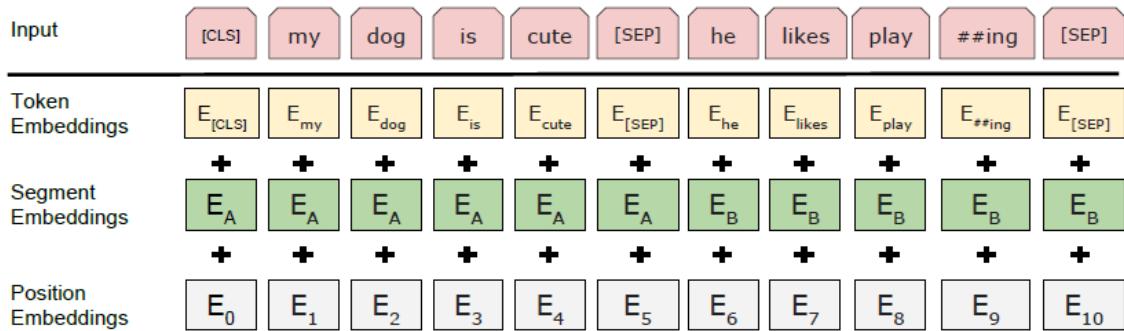
Dalam satu sekuens input, beberapa dari token input akan dilabeli [MASK] kemudian model akan memprediksi kalimat asli yang utuh dari token yang di-*mask* tersebut. Proses tersebut dilakukan terhadap token yang dipilih secara acak untuk melakukan *language modelling* secara *bidirectional*. Secara lebih spesifik, token yang dipilih untuk proses *masking* sebanyak 15%, kemudian 80% dari persentase tersebut kemudian di-*mask*, 10% diganti dengan kata yang dipilih secara random, dan 10% sisanya dilatih ulang.

2. *Next Sentence Prediction* (NSP)

Diberikan dua kalimat sebagai *sentence pair* A dan B. Sebagai contoh, terdapat sepasang kalimat yang berjumlah 100.000 pasang sebagai data latih, dengan skenario sebagai berikut :

- a. 50% dari *sentence pair*, kalimat keduanya adalah kalimat kedua yang sebenarnya dari kalimat pertama.
- b. 50% sisanya dari *sentence pair* tersebut, kalimat keduanya dipilih secara random dari korpus yang ada.
- c. Untuk kasus a akan dilabeli ‘*IsNext*’ sedangkan kasus b akan dilabeli ‘*NotNext*’.

Task ini digunakan untuk meng-*capture* relasi antara kalimat yang mana *language modelling* tidak dapat melakukan hal tersebut. Ilustrasi dari NSP dapat dilihat pada gambar II. 4.



Gambar II. 4 Ilustrasi NSP di mana *Input Embeddings* adalah jumlah dari *Token Embeddings*, *Segment Embeddings*, dan *Position Embeddings*

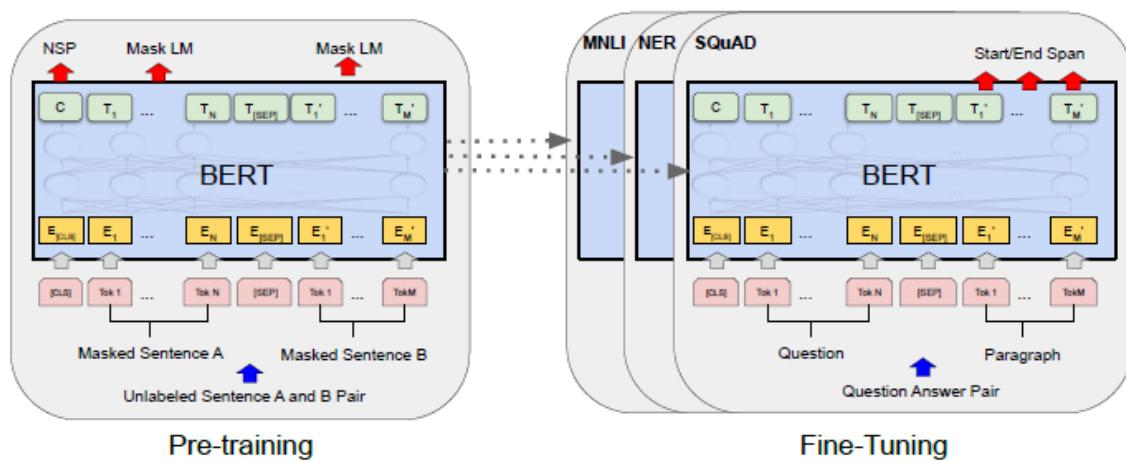
Kemampuan BERT dalam mengombinasikan MLM dengan NSP ini, yang menyebabkan BERT disebut sebagai model *task-agnostic* yang sesungguhnya.

Selanjutnya pada fase *fine-tuning*, secara sederhana cukup dengan menambahkan layer input dan output yang sesuai dengan task yang akan diselesaikan maka selanjutnya BERT akan melakukan *fine-tuning* secara *end-to-end* terhadap seluruh parameter. Sebagai contoh, jika terdapat kalimat masukan A dan B maka kalimat tersebut analog dengan *specific tasks* tertentu misalnya :

1. Pasangan kalimat untuk *paraphrasing task*.
2. Premis dan hipotesis untuk *entailment task*.
3. Pasangan *Question-passage* untuk kasus *Question Answering*.
4. Pasangan teks-∅ untuk klasifikasi teks dan *tagging* sekuensial.

Pada keluaran, representasi token dimasukkan ke layer *output* untuk *tasks* pada level token, *sequence tagging* dan *question answering*. Selain itu, representasi [CLS] dimasukkan ke layer output untuk kasus klasifikasi seperti *entailment* atau *sentiment analysis*. Saat ini, *encoder* pada kalimat atau dokumen yang menghasilkan representasi kontekstual token, telah di-*pre-train* pada data yang tidak beranotasi dan di-*fine-tune* untuk menyelesaikan kasus *supervised downstream*. Keuntungan dari pendekatan ini adalah parameter yang harus di-*train* menjadi lebih sedikit.

Pada BERT, tahap *pre-training* dengan *fine-tuning* menggunakan arsitektur yang sama. Parameter pada *pre-train* yang sama digunakan untuk menginisialisasi model untuk *downstream tasks* yang berbeda. Selama proses *fine-tuning* seluruh parameter di-*fine-tune*. Ilustrasi mekanisme *pre-training* dan *fine-tuning* dapat dilihat pada gambar II.5.



Gambar II.5 Mekanisme *pre-training* dan *fine-tuning*

[CLS] adalah simbol khusus yang ditambahkan di depan setiap kalimat masukan sedangkan [SEP] adalah *separator* token khusus misalnya untuk memisahkan *question/answer*.

II. 4 Metadata Linguistik

Penelitian sebelumnya telah menunjukkan bahwa kondisi mental seseorang mempengaruhi bahasa yang digunakan oleh penderita depresi (Trotzek et al., 2020). Oleh karena itu, seluruh fitur metadata berbasis teks akan diekstrak dari masing masing teks. Kemudian seluruh fitur akan dihitung secara terpisah dengan mencari nilai rata-ratanya atau dengan menjumlahkan nilainya. Adapun fitur yang akan diekstrak adalah sebagai berikut :

1. Kata dan Penggunaan Gramatikal

Sebagaimana telah disebutkan sebelumnya, bahwa depresi dapat mempengaruhi penggunaan kata dan tata bahasa yang digunakan oleh penderitanya. Salah satunya, adalah penderita depresi akan banyak menggunakan *personal pronouns*, *possessive pronouns*, serta bentuk kata kerja terdahulu (*past tense*) (Rude et al., 2004; Nadeem et al., 2016). Kata tersebut akan dijumlahkan kemunculannya dari teks masukan. Selain itu juga dihitung kemunculan kata spesifik seperti “depresi”, “kekhawatiran”, “terdiagnosa depresi”, serta antidepresan seperti Zoloft dan Paxil . Fitur berbasis leksikal tersebut didapat dari penelitian Oyong et al., (2018).

2. Keterbacaan

Nilai kompleksitas dari kalimat yang ditulis oleh penderita depresi menunjukkan perbedaan yang cukup signifikan jika dibandingkan dengan seseorang yang tidak terindikasi depresi. Untuk penderita yang tidak terindikasi depresi, lebih cenderung menulis dengan sederhana, bahkan dengan kalimat yang pendek (Trotzek et al., 2017). Beberapa *readability* matriks yang digunakan pada penelitian Trotzek et al. (2020) yakni *Gunning Fog Index* (FOG), *Flesch Reading Ease* (FRE), *Linsear Write Formula* (LWF), dan *New Dale-Chall Readability* (DCR) menunjukkan bahwa penderita depresi cenderung memiliki nilai *readability* yang lebih rendah jika dibandingkan dengan *control group* (tidak terindikasi depresi). Pada penelitian ini empat matriks *readability* yang akan digunakan untuk kasus Bahasa Indonesia adalah *Gunning Fog Index* (FOG), SMOG, *Flesch-Kincaid*, dan Dwiyanto Djoko Pranowo.

3. Polaritas Negatif

Emosi dari penderita menjadi salah satu fitur yang penting untuk mendeteksi depresi lebih khusus lagi, penderita depresi akan seringkali mengeluarkan *statement* ataupun tulisan, yang memiliki emosi yang negatif. Polaritas ini didapat dari kemunculan kata – kata terkait depresi dari penelitian Oyong et

al., (2018), yang mana kata tersebut diekstrak dari DSM-IV dan CES-D. Kata – kata tersebut terdiri dari 33 kata yang akan menjadi *seed* yang akan dicari sinonim tingkat satu dan tingkat dua menggunakan Kateglo API. Jumlah dari masing - masing kata tersebut yakni 187 kata dan 1.333 kata. Kata-kata tersebut kemudian di validasi oleh expert yakni seorang psikolog yang menilai apakah kata-kata tersebut sesuai dengan *diagnostic manual*.

4. Ringkasan Fitur Metadata

Fitur metadata yang digunakan pada penelitian ini mengacu pada penelitian Trotzek et al., (2020) dan Kumar et al., (2019), serta daftar kata yang menunjukkan indikasi depresi menggunakan data dari penelitian Oyong et al., (2018). Adapun ringkasan dari fitur metadata linguistik level pengguna pada penelitian ini dapat dilihat pada tabel II.1.

Fitur	Tipe	Deskripsi
Kata tunjuk orang pertama tunggal	Rata-rata	Berdasarkan POS <i>tagging</i>
Kepemilikan orang pertama	Rata-rata	Berdasarkan POS <i>tagging</i>
Waktu lampau	Rata-rata	Berdasarkan POS <i>tagging</i>
Nilai Keterbacaan	Rata-rata	<i>Gunning Fog Index (FOG)</i> , SMOG, <i>Flesch-Kincaid</i> , dan Dwiyanto Djoko Pranowo
Polaritas Negatif	Jumlah	-
Depresi	Jumlah	“Saya Depresi”
Diagnosis	Jumlah	“Saya di diagnosis depresi”
Antidepresan	Jumlah	“Zoloft”, “Paxil”

Tabel II.1 Ringkasan Fitur Metadata Linguistik Level Pengguna.

II. 5 Penelitian Terkait

II. 5. 1 CLEF 2017 (Losada & Crestani, 2016)

Penelitian yang dilakukan oleh Losada & Crestani (2016) ini bertujuan untuk melakukan diskusi dalam rangka pembuatan *benchmarks* untuk algoritma deteksi dini dengan mengevaluasi metodologi, efektivitas matriks dan proses lainnya yang terkait dengan deteksi dini. Hingga saat ini penelitian terkait depresi masih mengabaikan *temporal dimension* informasi tersebut penting dalam rangka menelusuri pola emosi dan mental seseorang. Oleh karena itu, Losada & Crestani (2016) membangun data set yang memperhatikan kronologis penulisan data. untuk data pada *stage training* dan *testing* dibagi menjadi masing - masing 10 *chunks* data yang terurut berdasarkan tanggal penulisan.

Sumber data yang digunakan pada CLEF eRisks 2017 dipilih berdasarkan ukuran dan kualitas sumber data, ketersediaan *history* dari interaksi individu dalam koleksi, kesulitan untuk membedakan depresi dan non-depresi serta syarat dan ketentuan dalam me-redistribusi data. Sehingga, Reddit dipilih sebagai sumber data pada penelitian ini karena memenuhi persyaratan diatas.

Dalam melakukan pelabelan depresi dan non-depresi dilakukan dengan melihat pernyataan bahwa pengguna yang bersangkutan didiagnosa depresi. Jika pernyataan didiagnosa depresi tersebut tersedia, maka pengguna yang bersangkutan akan dilabeli sebagai depresi. Adapun yang tidak menyatakan bahwa didiagnosa depresi akan dilabeli “*control*” atau tidak depresi termasuk dengan pernyataan depresi yang terindikasi adalah *self-diagnose*. Adapun teks yang diekstrak lebih detailnya adalah seperti pada tabel 1.

Tabel 1 Distribusi Data CLEF eRisks 2017

	<i>Depressed</i>	<i>Control</i>
Jumlah Subjek	137	755

Jumlah submission (<i>posts & comments</i>)	49.580	481.873
Rata-rata jumlah submission/subjek	361,9	638,2
Rata-rata jumlah kata / <i>submission</i>	27,4	36,7

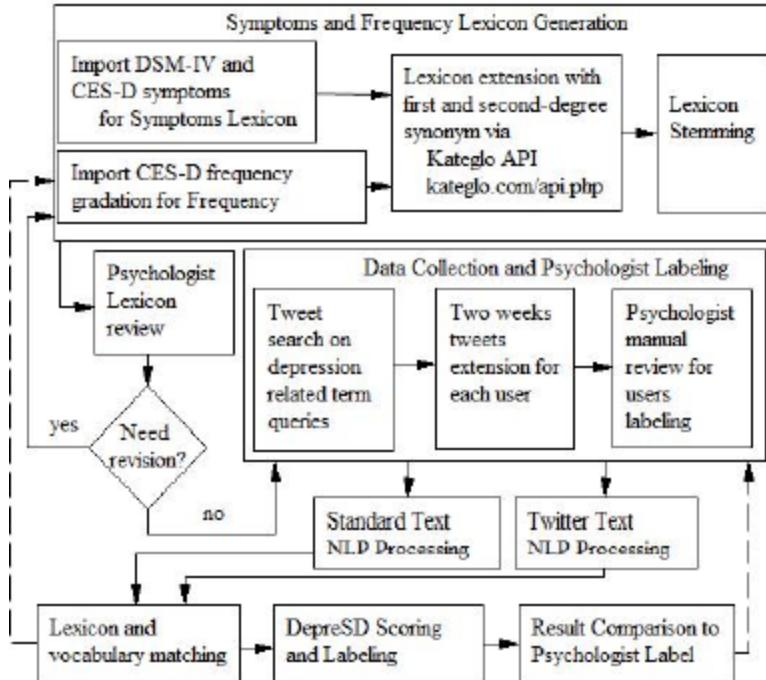
II. 5. 2 Oyong et al., (2018)

Penelitian yang dilakukan oleh Oyong et al., (2018) melakukan deteksi depresi menggunakan *tweet* berbahasa Indonesia dengan metode berbasis leksikal. Daftar kata yang menunjukkan gejala depresi diekstrak dari DSM-IV dan CES-D yang sudah divalidasi oleh psikologi. Gejala tersebut dicari sinonim katanya menggunakan Kateglo API sebuah *tool* tesaurus untuk Bahasa Indonesia. Kemudian dilakukan dua skema preprocessing yakni *Preprocessing Standar* dan *Preprocessing Khusus Teks Twitter*

Selanjutnya, dilakukan penghitungan skor depresi dengan formula yang diadopsi dari Christian Karmen, Robert H. Hsiung, dan Thomas Wetter pada tahun 2015, seperti berikut ini:

$$DepreSD = w_{pronouns} \cdot w_{negation} \cdot f_{frequencies} \cdot f_{symptoms} \quad (1)$$

Adapun arsitektur dari penelitian ini adalah seperti pada gambar II.6:



Gambar II.5 Arsitektur Penelitian Oyong et al., (2018)

II. 5. 3 (Kumar et al., 2019)

Penelitian yang dilakukan Kumar et al., (2019) tidak hanya melakukan deteksi kecemasan dan depresi berdasarkan gejala yang diukur melalui kemunculan kata terkait depresi tetapi juga menggunakan fitur-fitur lain, yakni 5-tuple vector seperti sebagai berikut :

a. *Anxiety Lexicon Base <w:word>*

Dibuat 60 list keyword yang merepresentasikan kecemasan dalam konten tekstual. Token yang telah diproses menggunakan WordNet ini dicocokan dengan leksikon dengan nilai fitur true (1) jika kata muncul dalam basis leksikon dan jika tidak diberi nilai 0.

b. *Tweets Timing: <t:timing>*

Untuk setiap tweet yang ditulis pada rentang pukul 12 malam hingga pukul enam pagi mengindikasikan pengguna akun memiliki gangguan tidur dan overthinking dilabeli 1, selain itu maka akan dilabeli 0.

d. Frekuensi: <f:frequency>

Fitur akan diatur 1 jika tweet di post sebanyak lebih dari 3 kali. Sebaliknya dilabeli 0.

e. *Negative Polarity*: <s:sentiment>

Masing – masing kata akan diubah menjadi synset dan di assign ke dalam 3 nilai antara 0.0 hingga 1.0 yang merepresentasikan positif, negative dan netral. Kata dengan rata-rata nilainya negatif selama 30 hari akan dijumlahkan. Jika nilai polaritas > 25% maka akan di *assign* nilai 1 sebaliknya akan di *assign* nilai 0.

f. *Polarity Contrass*: <c:contrast>

Untuk menghitung kontras dari polaritas digunakan formula seperti persamaan (2) :

$$c = \frac{(\delta \cdot PP + pw) - (\delta \cdot NP + nw)}{(\delta \cdot PP + pw) + (\delta \cdot NP + nw)} \quad (2)$$

Dimana,

pw adalah jumlah kata dengan polaritas opini positif

nw adalah jumlah kata dengan polaritas opini negatif

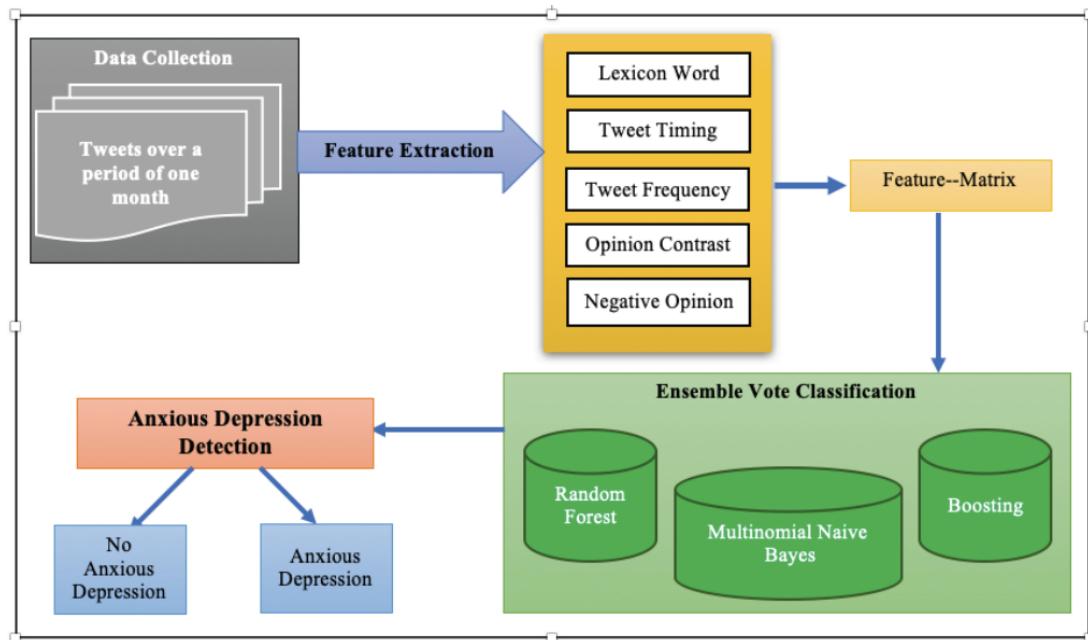
PP adalah jumlah postingan positif

NP adalah jumlah postingan negative

δ adalah koefisien postingan yang memiliki nilai 3

Jika magnitude dari *polarity contras* ini $\geq 25\%$ maka akan diberi nilai 1 sebaliknya akan diberi nilai 0.

Adapun arsitektur dari penelitian ini dapat dilihat pada gambar II. 6.



Gambar II.6 Arsitektur pada Penelitian Kumar et al., (2019)

Hasil dari penelitian tersebut adalah mendekati 85% prediksi telah terbukti akurat.

II. 5. 4 Trotzek et al., (2020)

Penelitian yang dilakukan oleh Trotzek et al., (2020) telah menunjukkan sudut pandang baru yang lebih spesifik dalam mendeteksi depresi. Trotzek et al., (2020) melakukan deteksi depresi dengan mengaplikasikan *Convolutional Neural Network* (CNN) dengan menggunakan beberapa *word embedding* yang berbeda yang kemudian akan dievaluasi dan dibandingkan performa masing-masing model klasifikasinya dengan model yang menggunakan metadata linguistik pada level pengguna. Untuk model *embedding* yang digunakan dapat dilihat pada gambar II. 7.

Model	Dimension	Corpus tokens (in billion)	Word vectors (in million)	Words of eRisk 2017
GloVe Wiki + News	50	6	0.4	81.9%
GloVe Crawl	300	42	2	93.2%
fastText Wiki	300	?	2.5	81.7%
fastText Wiki + News	300	16	1	79.5%
fastText Crawl	300	600	2	88.5%
fastText reddit	300	49.9	6	99.7%

Gambar II. 7 Model *Word Embedding* (Trotzek et al., 2020)

Pada penelitian ini, Trotzek et al., (2020) mencoba mengadopsi pendekatan yang dilakukan pada penelitian sebelumnya yakni pada eRisk 2017 (Trotzek et al., 2017) dimana fitur metadata linguistik menjadi masukan pada *fully-connected layer*, yang kemudian digabungkan dengan hasil dari layer LSTM. Namun, model tersebut tidak menunjukkan hasil yang lebih baik jika dibandingkan dengan hanya menggunakan teks sebagai input. Hal yang sama juga akan dilakukan dimana idenya adalah menggunakan vector dengan dimensi 1×50 sebelum layer *softmax* sebagai input tambahan untuk *classifier* metadata seperti *logistic regression*. Penelitian ini kemudian akan membandingkan antara hasil dari penelitian eRisk 2017 yang menggunakan nilai $ERDE_5$, $ERDE_{50}$, dan F_1 , dengan model yang diusulkan dengan menggunakan nilai $ERDE_{20\%}$, $ERDE_{50\%}$, dan $F_{latency}$.

Eksperimen dari penelitian ini, dilakukan dengan proses yang sama seperti pada penelitian eRisk 2017, yang mana data dibagi menjadi 10 bagian, kemudian proses *training* dilakukan pada seluruh *training dataset*. Selanjutnya, masing-masing bagian tadi diproses secara sekuensial terurut, yang mana data sebelumnya akan selalu digunakan bedanya adalah pada penelitian ini yang disebut dengan “*Meta LR Wait*”, yang mana *Logistic Regression* dikonfigurasi untuk melakukan *submit* terhadap hasil prediksi yang didapat hanya setelah bagian terakhir dari data yang

telah dibagi 10 tadi selesai diproses. Model dengan mekanisme *Meta LR Wait* ini berhasil mencapai nilai F_1 terbaik.

BAB III Analisis Masalah dan Rancangan Solusi

III. 1 Analisis Masalah

Berdasarkan penelitian sebelumnya oleh Oyong et al., (2018) yang dilakukan dengan pendekatan berbasis leksikal. Oyong et al., (2018) menghitung bobot gejala depresi berdasarkan kemunculan daftar gejala depresi yang sudah diekstrak berdasarkan DSM-IV dan CES-D. Permasalahan yang terjadi diantaranya adalah sebagai berikut:

1. Kesalahan klasifikasi yang disebabkan terjadinya OOV (*out-of-vocabulary*) disebabkan karena metode yang digunakan berbasis pembobotan gejala dari kata gejala yang muncul pada data pengguna twitter, jadi dapat diperkirakan bahwa bisa jadi ada seorang pengguna non depresi yang banyak membicara tentang depresi tetapi yang dibicarakan adalah tentang pengalaman keluarga atau temannya yang pernah mengalami depresi. Untuk itu butuh fitur lain yang dapat membantu meningkatkan akurasi dari model yang akan dibangun.
2. Masalah misklasifikasi hasil prediksi yang terjadi, dapat disebabkan karena jumlah bobot depresi untuk pengguna yang tidak depresi yang cukup besaran melewati *threshold* yang sudah ditentukan, sehingga model sebelumnya yang dibangun oleh Oyong et al., (2018) akan memprediksi pengguna tersebut sebagai penderita depresi dan begitu juga sebaliknya. Padahal dalam menentukan prediks, konteks kalimat juga harus dipahami oleh karena itu, pada penelitian ini akan menggunakan model bahasa yang akan memperhatikan konteks dari kalimat.
3. Dalam melakukan justifikasi penderita depresi pada penelitian sebelumnya dilihat berdasarkan jumlah kemunculan *term* yang terkait dengan gejala depresi. Jika sudah melewati nilai ambang tertentu, maka objek penelitian tersebut akan dilabeli sebagai depresi. Namun, dalam mendeteksi depresi,

perlu dipertimbangkan kondisi emosi seseorang dalam waktu tertentu. Hal ini dikarenakan kondisi seseorang yang dapat berubah-ubah secara signifikan. Oleh karena itu, menyatakan seseorang depresi atau tidak depresi tanpa melihat data secara historikal masih dinilai kurang tepat.

III. 2 Analisis Solusi

Berdasarkan masalah yang sudah didefinisikan diatas maka solusi yang akan diterapkan untuk mengatasi masalah tersebut adalah sebagai berikut:

1. Untuk mengatasi kesalahan klasifikasi akibat OOV, maka perlu ditambahkan fitur-fitur baru yang dapat meningkatkan akurasi dari model, yang kemudian disebut dengan fitur metadata linguistik. Adapun fitur metadata linguistik yang digunakan pada penelitian ini adalah sebagai berikut:
 - a. Daftar Kata Terindikasi Depresi

Daftar kata ini diekstrak pada penelitian sebelumnya, yakni oleh Oyong et al.,(2018) yang telah mengumpulkan yang terdiri dari 34 kata. Daftar kata ini diekstrak dari dokumen DSM IV dan CES-D yang kemudian sudah divalidasi oleh psikologi.

- b. *Depressive Word*

Pada penelitian ini, beberapa kata yang terkait gejala depresi akan diberi bobot lebih, yang mana beberapa kata tersebut mengikuti daftar kata yang diusulkan pada penelitian Trotzek et al., (2017) yakni Depresi, Diagnosa, Mati, Bunuh Diri. Berdasarkan pengamatan pada data keempat kata ini termasuk kata-kata yang sering muncul pada pengguna yang dilabeli depresi.

- c. *Readability Score*

Pada penelitian sebelumnya, ditemukan bahwa penderita depresi lebih sulit mengungkapkan apa yang ada di pikirannya sehingga, seringkali kata-kata

yang diucapkan atau ditulis lebih sulit untuk dipahami *readability score* menghitung seberapa mudah suatu kalimat dapat dipahami. Adapun algoritma yang akan digunakan untuk menghitung nilai keterbacaan ini adalah algoritma *gunning fog*. Adapun langkah yang digunakan untuk melakukan penghitungan keterbacaan dengan algoritma *gunning fog* adalah sebagai berikut:

- i. Ambil 100 kata dari masing-masing teks.
- ii. Lakukan pembagian antara jumlah kata dibagi dengan jumlah kalimat.
- iii. Hitung jumlah kata kompleks yakni kata dengan jumlah suku kata lebih dari 3.
- iv. Lakukan pembagian antara kata kompleks dengan jumlah total kata.
- v. Kalikan hasil pembagian dari langkah iv dengan 100 kemudian tambahkan hasil penjumlahan pada nomor ii.
- vi. Kalikan hasil dari langkah 5 dengan 0,4.

Adapun rumus untuk mencari nilai *Gunning Fog* sebagaimana dapat dilihat pada formula 1.

$$\text{Nilai index} = 0,4 \left(\frac{\text{Jumlah kata}}{\text{Jumlah kalimat}} \right) + 100 \left(\frac{\text{Jumlah kata kompleks}}{\text{Jumlah kata}} \right) \quad (1)$$

Untuk menerapkan algoritma *Gunning Fog*, terdapat aturan yang harus diikuti, yakni adalah sebagai berikut:

- a. Pada akhir kalimat, harus ditandai dengan tanda titik (.), tanda tanya (?) dan tanda seru (!). Bukan tanda koma (,), titik dua (:) maupun titik koma (;).
- b. Jangan menghitung kata majemuk yang memiliki tanda penghubung.
- c. Hitung jumlah suku kata pada setiap kata yang merupakan kata keras.

- d. Hitung kata singkatan dengan melihat kata aslinya.
- e. Daftar dihitung sebagai satu kalimat baik yang dipisah dengan titik koma, maupun dengan koma.

Oleh karena itu pada penelitian ini, akan digunakan fitur-fitur tersebut untuk melakukan deteksi depresi yang disebut dengan fitur metadata linguistik

2. Untuk mengatasi masalah misklasifikasi akibat ketidakmampuan model sebelumnya dalam memahami konteks dari kalimat, sehingga dibutuhkan model yang mampu semantik di dalam kalimat. Hal ini dapat dilakukan dengan memanfaatkan *cross lingual transfer learning* dengan XLM-ROBERTa yang mana saat ini sudah mencapai *state-of-the-art* dan juga M-BERT. Pada penelitian ini, model tidak hanya harus mampu memahami konteks kalimat, namun juga mampu melakukan pembelajaran transfer dari *task* berbahasa Inggris yakni dengan data CLEF 2017 ke *task* berbahasa Indonesia. Pada penelitian ini XLM-R dan M-BERT akan digunakan sebagai model pembelajaran transfer berbasis konteks dalam menyelesaikan masalah klasifikasi.
3. Permasalahan yang lain yaitu dalam melakukan justifikasi penderita depresi, harus melihat data secara historikal, dan melihat bagaimana pola emosi objek penelitian. Hal ini dapat dilakukan dengan menggunakan ERDE (*Early Risk Detection Error*) yang tidak hanya dapat menentukan objek penelitian adalah depresi atau tidak, tetapi juga dapat menjadi salah satu alat untuk melakukan evaluasi terkait model yang dibangun.

III. 3 Desain Eksperimen

Terdapat dua eksperimen yang dilakukan pada penelitian ini, yakni yang pertama adalah penelitian ini akan melihat bagaimana performa model yang dibangun dengan menggunakan *pre-train* model bahasa XLM-RoBERTa dan BERT Multilingual yang akan dikombinasikan dengan metadata linguistik pada level user yang akan dilakukan *pre-train* dengan menggunakan data berbahasa Inggris dari

CLEF eRisks 2017. Hal ini dilakukan untuk mengatasi OOV dan kesalahan klasifikasi yang dilakukan pada model penelitian berikutnya. Kemudian eksperimen kedua adalah melakukan pencarian kombinasi parameter yang akan digunakan.

III. 4 Metode Evaluasi Kinerja

Metode evaluasi kinerja yang digunakan pada penelitian ini menggunakan mekanisme evaluasi yang dilakukan oleh David yang mengusulkan *Early Risk Detection Error* (ERDE) yang melakukan evaluasi dan melakukan justifikasi terhadap penderita depresi dari serangkaian label yang dihasilkan dari model yang dibangun.

Untuk masing-masing pengguna diberikan 1 sekuens data. Mesin dapat memberikan label salah satu dari depresi, kontrol atau model juga dapat memberikan *delay* jika model belum dapat menentukan label. *Delay* ini kemudian akan dihitung berdasarkan data yang dilihat sampai keputusan dapat dibuat oleh model. Sebagai contoh jika terdapat pengguna p memiliki 25 kalimat atau postingan setiap *chunk* datanya, maka jika model dapat memberikan keputusan setelah *chunk* kedua maka nilai penalti yang diberikan adalah 50 model butuh 50 bukti untuk menentukan keputusan. Namun, kenyataannya masalah data *unbalanced* seringkali terjadi. Untuk itu model ini memberikan bobot berdasarkan kasus eror yang berbeda, ERDE mendefinisikan perhitungannya adalah sebagai berikut:

Untuk nilai c_{fp} dan c_{fn} tergantung pada domain aplikasi dan implikasi dari FP dan FN. Untuk kasus di mana data negatif lebih banyak dari pada data positif , maka model akan lebih sering menyatakan negatif, sehingga untuk mencegah hal tersebut nilai c_{fn} harus lebih besar c_{fp} dengan cara memberikan nilai 1 pada c_{fp} dan memberikan nilai c_{fn} yang sesuai dengan proporsi data positif yang dimiliki.

Sedangkan faktor $lc_o(k)$ ($\in [0, 1]$) melakukan *encode* yang berkaitan dengan *delay* dalam mendeteksi *true positive*. Fungsi $lc_o(k)$ adalah fungsi yang meningkatkan fungsi k atau *delay*:

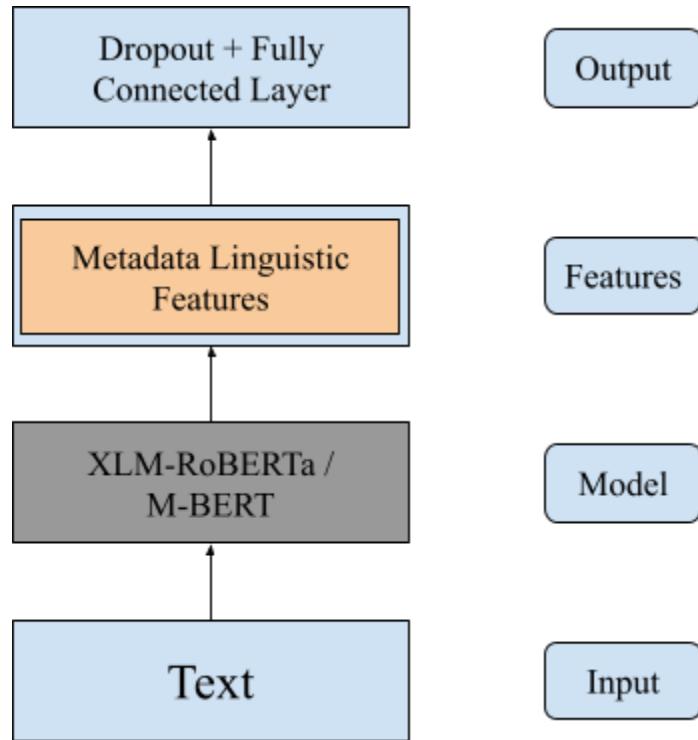
$$lc_o(k) = 1 - \frac{1}{1 + e^{k-o}} \quad (2)$$

III. 5 Rancangan Solusi

Berdasarkan analisis yang dilakukan pada subbab III.1 dan III.2 maka arsitektur penelitian yang akan digunakan adalah sebagaimana pada gambar III.1. Arsitektur ini secara umum mengikuti arsitektur yang dilakukan pada penelitian Trotzek, namun pada penelitian ini akan memanfaatkan model *cross-lingual* yang akan melakukan *pre-train* berbahasa Inggris dengan data CLEF eRisks 2017 yang nantinya akan digunakan untuk memprediksi depresi berbahasa Indonesia. Mula-mula data latih akan di tokenisasi misalnya untuk kalimat “*I feel so unhappy because there is no one core reason.*“ maka hasil tokenisasi adalah sebagai berikut:

```
['_I', '_feel', '_so', '_un', 'happy', '_because',
'_there', '_is', '_no', '_one', '_core', '_reason', '.']
```

proses selanjutnya adalah melakukan *encoding* terhadap teks yang sudah di tokenisasi untuk menghasilkan representasi teks kontekstual untuk selanjutnya masuk ke metadata linguistik *layer* dengan dimensi 1×50 sebagai input. Selanjutnya hasil prediksi akan dikeluarkan melalui layer output yang merupakan *fully connected layer*.



Gambar III.1 Arsitektur penelitian

DAFTAR PUSTAKA

- Azhar, A. N. (2020). *PEMBELAJARAN TRANSFER UNTUK BAHASA INDONESIA MENGGUNAKAN PENDEKATAN FINE-TUNING MODEL BAHASA*. 82.
- Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.
- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2015). *On the Properties of Neural Machine Translation: Encoder–Decoder Approaches*. 103–111. <https://doi.org/10.3115/v1/w14-4012>
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. *ArXiv*. <https://doi.org/10.18653/v1/2020.acl-main.747>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), 4171–4186.
- El-Alami, F. zahra, Ouatik El Alaoui, S., & En Nahnhahi, N. (2021). Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization. *Journal of King Saud University - Computer and Information Sciences*, xxxx. <https://doi.org/10.1016/j.jksuci.2021.02.005>
- Ikhwantri, F. (2019). Cross-Lingual Transfer for Distantly Supervised and Low-resources Indonesian NER. *ArXiv*.
- Joshi, D., & Patwardhan, D. M. (2020). An analysis of mental health of social

- media users using unsupervised approach. *Computers in Human Behavior Reports*, 2(August), 100036. <https://doi.org/10.1016/j.chbr.2020.100036>
- Kumar, A., Sharma, A., & Arora, A. (2019). Anxious Depression Prediction in Real-time Social Data. *International Conference on Advanced Engineering, Science, Management and Technology – 2019 (ICAESMT19)*, 2019, 1–7.
- Losada D.E., Crestani F. (2016) A Test Collection for Research on Depression and Language Use. In: Fuhr N. et al. (eds) Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2016. Lecture Notes in Computer Science, vol 9822. Springer, Cham.
https://doi.org/10.1007/978-3-319-44564-9_3
- Nayak, A., Timmapathini, H., Ponnalagu, K., & Gopalan Venkoparao, V. (2020). *Domain adaptation challenges of BERT in tokenization and sub-word representations of Out-of-Vocabulary words*. 1–5.
<https://doi.org/10.18653/v1/2020.insights-1.1>
- Oyong, I., Utami, E., & Luthfi, E. T. (2018). Natural language processing and lexical approach for depression symptoms screening of Indonesian twitter user. *Proceedings of 2018 10th International Conference on Information Technology and Electrical Engineering: Smart Technology for Better Society, ICITEE 2018*, 359–364. <https://doi.org/10.1109/ICITEED.2018.8534929>
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representation. *Naacl, Proceeding*, 2227–2237.
- Ruder, S. (2019). *Neural Transfer Learning for Natural Language Processing*.
- Sisk, D. (2018). Deep Learning for Depression Detection of Twitter Users. *Gifted Child Quarterly*, 19(2), 175–180.
<https://doi.org/10.1177/001698627501900225>
- Tai, C. M. Y., Theune, M., Seifert, C., & Möller, T. (2019). Effects of Inserting

- Domain Vocabulary and Fine-tuning BERT for German Legal Language.
Master's Thesis.
https://essay.utwente.nl/80128/1/Yeung_InteractionTechnology_EEMCS.pdf
- Trotzek, M., Koitka, S., & Friedrich, C. M. (2020). Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences. *IEEE Transactions on Knowledge and Data Engineering*, 32(3), 588–601. <https://doi.org/10.1109/TKDE.2018.2885515>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kalser, L., & Polosukhin, I. (2017). *Attention Is All You Need. Nips*.
- Wang, H., Yu, D., Sun, K., Chen, J., & Yu, D. (2019). Improving pre-trained multilingual models with vocabulary expansion. *ArXiv*, 316–327.
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and resources for evaluating indonesian natural language understanding. *ArXiv*.
- Yang, Z., Salakhutdinov, R., & Cohen, W. W. (2017). Transfer learning for sequence tagging with hierarchical recurrent networks. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 1–10.