

PENGENALAN ENTITAS BIOMEDIS DALAM TEKS KONSULTASI KESEHATAN ONLINE BERBAHASA INDONESIA BERBASIS ARSITEKTUR TRANSFORMERS

Abid Famasya Abdillah¹, Diana Purwitasari^{*2}, Safitri Junita³, Mauridhi Hery Purnomo⁴

^{1,2,3,4} Institut Teknologi Sepuluh Nopember, Surabaya, ³Universitas Budi Luhur, Jakarta
Email: ¹abid.famasya@gmail.com, ²diana@if.its.ac.id, ³safitri.juanita@budiluhur.ac.id, ⁴hery@ee.its.ac.id
^{*}penulis korespondensi

(Naskah masuk: 09 Juni 2022, diterima untuk diterbitkan: 28 Februari 2023)

Abstrak

Pengenalan entitas biomedis merupakan salah satu tahapan penting dalam ekstraksi informasi pada domain kesehatan. Untuk melakukannya, penelitian terkini banyak menggunakan model ekstraksi biomedis berbasis *deep learning* yang juga dikenal sebagai *Biomedical NER* (BioNER). Banyak penelitian menggunakan data sosial media sebagai basis data latih BioNER untuk memenuhi kebutuhan data yang besar. Di sisi lain, banyaknya topik bahasan pada sosial media membuat sumber data ini kurang representatif digunakan dalam pelatihan BioNER seiring dengan melimpahnya bias serta kurangnya data terkait biomedis. Oleh karena itu, penelitian ini mengusulkan suatu model BioNER yang telah dilatih pada situs konsultasi kesehatan *online* (KKO) agar memiliki representasi data medis lebih baik dibandingkan dengan penelitian lain yang sejenis. Kontribusi utama penelitian ini adalah terbentuknya model BioNER yang dapat digunakan dalam metode ekstraksi informasi biomedis dalam Bahasa Indonesia. Model ini dibangun menggunakan arsitektur *state-of-the-art* Transformers sehingga mendapatkan hasil evaluasi F1 *score* sebesar 0.7691, mengungguli model LSTM sebesar 0.03 poin. Hasil simulasi terhadap data riil juga menunjukkan bahwa model BioNER mampu mengenali entitas biomedis secara umum meskipun dilatih pada data yang terbatas. Selain itu, dengan digunakannya model berbasis XLM-R, maka model juga memiliki kemampuan pengenalan multibahasa sehingga potensi implementasinya tidak terbatas pada entitas Bahasa Indonesia saja. Untuk mendukung penelitian lanjutan, model pengenalan entitas biomedis ini juga dapat diakses secara publik untuk di <https://huggingface.co/abid/indonesia-bioner>.

Kata kunci: pengenalan entitas biomedis, arsitektur transformers, konsultasi kesehatan online

BIOMEDIAL ENTITY RECOGNITION IN INDONESIAN ONLINE HEALTH CONSULTATION BASED ON TRANSFORMERS ARCHITECTURE

Abstract

Biomedical entity recognition is one of the important stage in the information extraction, particularly in the health domain. Recent research uses a deep learning-based biomedical extraction model known as Biomedical NER (BioNER). Due to extensive data requirement, many studies still use social media data as a BioNER training data. On the other hand, social media data is less representative because it contains a lot of bias and lack of medical representation terms as the impact of many topics discussed. Therefore, this study proposes a BioNER model that has trained on an online health consultation platform to gain a better representation of biomedical data. This model also built using the state-of-the-art Transformers architecture. Hence, its evaluation results show that this model is able to achieve an F1 score of 0.7691, outperforming the LSTM model by 0.03. Simulation results on the real data also indicate that the BioNER model is able to recognize biomedical entities in general cases despite only trained on limited data. In addition, by using an XLM-R-based model, the recognition model also has multilingual recognition capabilities. Therefore, there is a potential implementation to apply the our BioNER model beyond Indonesian biomedical entities. Our biomedical entity recognition model is also accessible at <https://huggingface.co/abid/indonesia-bioner>.

Keywords: biomedical entity recognition, transformers architecture, online health consultation

1. PENDAHULUAN

Masifnya pertumbuhan pengguna internet di Indonesia turut serta mengubah pola perilaku

masyarakat dalam mengakses informasi kesehatan. Pembatasan kegiatan masyarakat karena pandemi juga mengakselerasi penggunaan *online platform*

seperti sosial media dan forum untuk bertanya dan berdiskusi tentang kondisi kesehatan yang dialami. Sebagai contoh, sosial media kerap dipakai untuk bercerita dan berdiskusi tentang kondisi kesehatan penggunanya. Pengungkapan informasi kesehatan individu seperti gejala penyakit, kelainan, dan keluhan dari unggahan pengguna tersebut memiliki potensi kegunaan untuk analisis kesehatan publik. Sehingga, banyak penelitian mulai menggunakan data *online platform* sebagai data masukan untuk sistem pengawasan kesehatan publik.

Untuk mendapatkan informasi kesehatan yang relevan dari data *online platform*, salah satu tahap yang terpenting adalah ekstraksi informasi biomedis. Penelitian tentang pengenalan penyakit ringan di Twitter menggunakan ekstraksi gejala penyakit dalam bahasa Inggris, Jepang dan China sebagai basisnya (Wakamiya, et al., 2019). Dengan menggunakan *ensemble method* dari CNN dan *hierarchical attention*, penelitian tersebut diklaim mampu mengklasifikasikan *tweet* yang dikumpulkan dalam periode tertentu. Penelitian tentang sistem pakar dalam konteks COVID-19 juga membutuhkan data lokasi dan kata kunci tertentu untuk dapat mengenali tren naiknya kasus wabah di Amerika Serikat (Zhang, et al., 2022). Dalam konteks permasalahan yang sama, deteksi emosi dengan metode *topic modeling* juga digunakan untuk menganalisis sentimen masyarakat terhadap vaksin (Lyu, et al., 2021). Penelitian-penelitian tersebut menunjukkan bahwa fase ekstraksi informasi biomedis memiliki peranan yang penting dalam pengembangan sistem pengawasan kesehatan publik.

Salah satu metode ekstraksi informasi biomedis yang umum digunakan adalah *Biomedical named entity recognition* (BioNER). Untuk menghasilkan *state-of-the-art* BioNER, penelitian terkini banyak menggunakan arsitektur model berbasis *deep learning* (Song, et al., 2021). Model *deep learning* juga dikenal membutuhkan data latih yang besar untuk mendapatkan performa terbaik, sehingga data sosial media banyak digunakan karena kemudahannya mendapatkannya. Namun, penggunaan data sosial media dinilai kurang optimal sebagai representasi data latih biomedis, terutama dikarenakan kestabilan dan validitas datanya (Aiello, et al., 2020). Pertama, data sosial media dinilai tidak stabil karena banyaknya topik bahasan mengakibatkan data yang diperoleh seringkali tidak fokus membahas tentang isu kesehatan. Ditambah lagi, data sosial media seringkali membutuhkan *keyword seeds* untuk menemukan kata kunci yang tepat dalam pencarian data biomedis. Dalam hal validitas data, isu yang disoroti adalah banyak *bot* dan *influencer* yang mempengaruhi konteks bahasan di masyarakat. Akibatnya, perlu evaluasi konteks yang teliti sebelum menggunakan data sosial media sebagai basis pembangunan BioNER.

Di sisi lain, data situs konsultasi kesehatan *online* (KKO) memiliki informasi kontekstual yang

lebih baik untuk merepresentasikan informasi biomedis (Arachchige, et al., 2021). Penelitian untuk mengolah informasi implisit dari unggahan pengguna di platform konsultasi kesehatan dapat digunakan untuk mengidentifikasi dan memetakan sebaran usia pasien (Black, et al., 2022). Penelitian lain tentang pemanfaatan forum konsultasi kesehatan untuk mengevaluasi topik bahasan juga menunjukkan bahwa data yang digunakan memiliki konteks yang jelas (Athira, et al., 2021). Secara umum, seluruh penelitian tersebut menunjukkan bahwa data situs KKO dapat merepresentasikan informasi kesehatan yang terjadi di masyarakat. Meski demikian eksplorasi penggunaan data ini masih terbatas, sehingga penelitian lanjutan perlu dilakukan untuk melihat potensi penggunaan data tersebut, khususnya dalam konteks ekstraksi informasi biomedis.

Dengan pentingnya representasi data yang tepat untuk membentuk model ekstraksi biomedis, penelitian ini bertujuan untuk mengevaluasi model pengenalan entitas biomedis dalam bahasa Indonesia yang dibangun berdasarkan data situs KKO. Sumber data ini yang juga membedakan penelitian ini dengan penelitian lain yang sejenis. Selain itu, penggunaan arsitektur Transformers yang merupakan *state-of-the-art* dari *deep learning*, diharapkan model yang dibangun memiliki performa yang *robust* dalam mengenali entitas biomedis. Kontribusi utama dari penelitian ini adalah terciptanya suatu model BioNER dalam bahasa Indonesia yang dilatih pada data KKO, sehingga dapat dimanfaatkan sebagai metode ekstraksi informasi khususnya pada penelitian teks domain kesehatan.

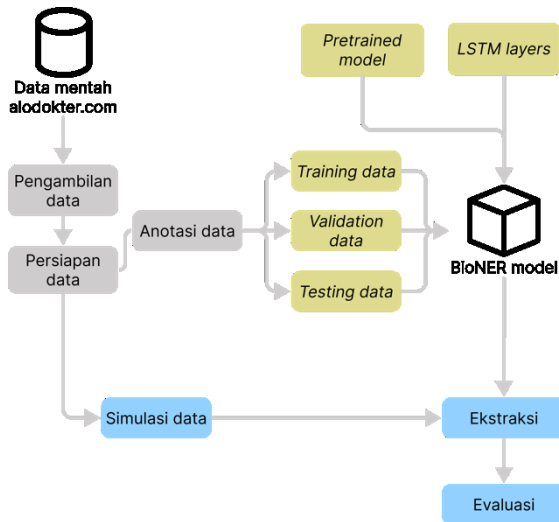
Secara garis besar penelitian ini akan dibagi menjadi lima bagian. Bagian pertama menjelaskan tentang latar belakang penelitian dilakukan, yang disusul dengan bagian kedua tentang pengenalan arsitektur Transformers. Bagian ketiga akan membahas metode penelitian dan skenario uji coba. Bagian keempat berisi diskusi hasil uji coba. Bagian terakhir adalah kesimpulan dan potensi pengembangan penelitian kedepannya.

2. METODE PENELITIAN

Untuk membangun BioNER, penelitian ini dibagi menjadi tiga bagian utama, yaitu pengumpulan dan persiapan data (abu-abu), pembangunan model (kuning) BioNER dan simulasi model (biru).

2.1. Pengumpulan dan persiapan data

Langkah pertama dari penelitian ini adalah pengumpulan data dari situs KKO. Banyak situs KKO yang dapat digunakan sebagai sumber data. Namun penelitian ini memilih Alodokter karena alasan-alasan berikut. Pertama, Alodokter merupakan situs yang paling aktif digunakan oleh pengguna.



Gambar 1. Metode pembangunan BioNER untuk identifikasi entitas biomedis dalam Bahasa Indonesia

Kedua, situs ini menggunakan dokter untuk menjawab pertanyaan yang datang sehingga jawaban yang diberikan terjaga validasinya. Pengambilan data dilakukan dengan teknik *web scraping* menggunakan pustaka pemrograman Scrapy¹ untuk setiap pertanyaan dan jawaban antara tahun 2014 hingga 2020. Hasilnya, didapatkan data dengan statistik sebagai berikut:

Tabel 1. Contoh label yang diberikan pada tiap kata di data anotasi

Properti	Nilai
Jumlah data	325.704
Jumlah kategori penyakit	1.004
Rata-rata kata / dokumen	55
Rata-rata karakter / dokumen	352

Setelah proses pengumpulan, 2.599 data dipilih dari empat kelompok penyakit untuk dilakukan anotasi entitas oleh dua mahasiswa kedokteran sebagai pakar dengan panduan dari *Unified Medical Language System* (UMLS) (Campillos-Llanos, et al., 2021). Empat kelompok tersebut adalah HIV, AIDS, diare dan TBC sebagai penyakit berisiko tinggi yang juga paling banyak ditanyakan pada situs Alodokter.

Tags diberikan dalam format IOB untuk menyimpan informasi posisi sebuah kata dalam suatu entitas. Sedangkan entitas yang diberikan adalah *disorders* (DISO) yang meliputi segala jenis gejala penyakit hingga kelainan tubuh, serta *anatomy* (ANAT) mencakup seluruh bagian tubuh hingga tingkat sel. Hasilnya, lima tags {B-DISO, I-DISO, B-ANAT, I-ANAT, O} telah diberikan untuk menandai 734.793 kata di seluruh data latih. Contoh data anotasi yang diberikan dapat dilihat pada Tabel 2.

Langkah selanjutnya adalah persiapan data. Pada tahapan ini, tidak dilakukan *stopwords removal* seperti yang dilakukan pada metode *machine learning* pada umumnya karena anotasi hanya diberikan kepada entitas yang penting pada dokumen.

Selain itu, suatu penelitian juga menunjukkan bahwa *stopwords* hanya berpengaruh sedikit terhadap performa model *deep learning* (Camacho-Collados & Pilehvar, 2018). Sehingga persiapan data pada tahap ini adalah menghilangkan karakter non-alfanumerik pada seluruh korpus.

Tabel 2. Contoh label yang diberikan pada tiap kata di data anotasi

Kata	Label
Kaki	B-ANAT
kiri	I-ANAT
saya	O
nyeri	B-DISO
sekali	I-DISO
kemarin	O
...	...

2.2. Pembangunan model BioNER

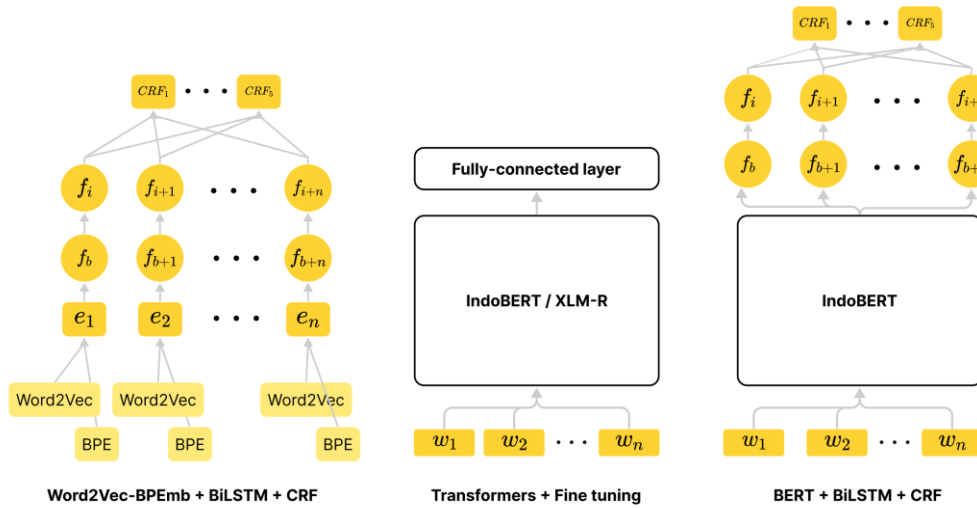
Pada awal tahap ini, data yang telah dianotasi pada tahap sebelumnya dibagi menjadi tiga bagian, yaitu 1.559 dokumen *data training*, 260 dokumen *data validation* dan 780 dokumen *data testing*. Terdapat tiga arsitektur yang akan diujicoba pada penelitian ini: LSTM, Transformers dan *hybrid* Transformers-LSTM. Arsitektur LSTM digunakan sebagai pembanding model karena banyak digunakan penelitian lain untuk membentuk model NER dan BioNER, sedangkan Transformers merupakan model usulan yang digunakan pada penelitian ini. Selain itu, ujicoba ini melibatkan data *hybrid* untuk mengevaluasi apakah data gabungan Transformers dan LSTM memiliki performa yang lebih baik dibandingkan arsitektur normalnya.

Pada arsitektur LSTM, penelitian ini menggunakan kombinasi antara Word2Vec dan BPEmb sebagai *word embedding* untuk mendapatkan nilai *embedding* yang juga mampu menangani permasalahan OOV (Heijden, et al., 2020). Di sisi lain, model *bidirectional* LSTM digunakan agar *information flow* antar *states* dalam LSTM dapat berjalan dua arah (*forward* dan *backward*). Di akhir model model, CRF layer ditempatkan berdasar evaluasi bahwa CRF mampu meningkatkan nilai akurasi dari model *deep learning* (Jin & Yu, 2021). Sehingga bentuk akhir dari model ini adalah Word2Vec-BPEmb + Bi-LSTM + CRF.

Untuk mengetahui parameter paling optimal pada arsitektur LSTM, *hyperparameters search* juga dilakukan dengan menggunakan Hyperopt (Bergstra, et al., 2015). Pencarian parameter yang akan dicari adalah sebagai berikut: *hidden_size* = [8, 16, 32, 64, 128]; *learning_rate* = [0.001 – 0.5]; *dropout_rate* = [0.00001 – 0.1]; *batch_size* = [8, 16, 32, 64, 128].

Untuk arsitektur Transformers, IndoBERT dan XLM-R digunakan sebagai *pretrained model*. IndoBERT merupakan versi dari BERT yang dilatih khusus pada sumber data berbahasa Indonesia seperti Wikipedia dan berita (Koto, et al., 2020).

¹ <https://scrapy.org/>



Gambar 2. Tiga arsitektur uji coba yang terdiri dari Word2Vec-BPEmb + BiLSTM + CRF, Transformers dan *fine tuning*, serta BERT + BiLSTM + CRF.

Sedangkan XLM-R adalah model berbasis Transformers yang dilatih pada sumber data dari berbagai bahasa, sehingga memiliki jumlah parameter yang lebih banyak daripada BERT dan memiliki kemampuan *encoding* data multibahasa (Conneau, et al., 2020). *Fine tuning* dilakukan dengan menambahkan *fully-connected layer* di akhir kedua *pretrained model* dengan output sebanyak lima kelas.

Model terakhir yang akan dibangun adalah model *hybrid* BERT + BiLSTM + CRF berdasarkan (Dai, et al., 2019). Model ini menggunakan IndoBERT sebagai *encoder* teks, dimana output dari *pretrained model* akan diteruskan ke *bidirectional LSTM* serta CRF layer. Tujuan dari pembangunan *hybrid model* ini untuk mengevaluasi apakah model yang lebih kompleks akan memiliki performa yang lebih baik dibanding *fine tuning* standar ataupun tidak.

2.3. Evaluasi dan simulasi model

Langkah terakhir dari tahapan penelitian adalah evaluasi performa dari model yang telah dibentuk sebelumnya dengan menggunakan *macro Precision*, *Recall* dan *F1 score*. *Macro score* dipilih karena seluruh kelas memiliki dianggap bobot yang setara. Nilai *F1 score* juga merupakan penentu diantara seluruh matriks karena nilai *F1* yang tinggi menunjukkan keseimbangan antara *Precision* dan *Recall*. Secara formal, *Precision*, *Recall* dan *F1 score* dihitung dalam Persamaan 2, 3 dan 4 berturut-turut.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

Karena satu entitas biomedis bisa terdiri dari beberapa kata sekaligus, maka terdapat beberapa jenis

perhitungan kebenaran entitas. Pada penelitian ini digunakan jenis perhitungan *relaxed match* karena lebih sesuai dengan kondisi pada dunia nyata (Tsai, et al., 2006). Pada perhitungan *relaxed match*, *prefix* dari label akan dihapuskan. Sebagai contoh, jika terdapat *predicted* dan *actual label* berikut:

$$\text{Prediction} = [B - \text{DISO}, B - \text{DISO}, I - \text{DISO}]$$

$$\text{Actual} = [B - \text{DISO}, I - \text{DISO}, I - \text{DISO}]$$

Maka nilai prediksi akan tetap mendapat nilai 100% dikarenakan *prefix* B dan I diabaikan.

Selain evaluasi, penelitian ini juga akan melakukan simulasi dengan data yang tidak terannotasi untuk melihat performa model melakukan ekstraksi entitas biomedis pada data riil. Data *sample* simulasi diambil dari situs Alodokter pada tahun 2021 untuk melihat bagaimana model melakukan ekstraksi informasi diluar kategori penyakit yang menjadi data latih sebelumnya. Melalui ekstraksi, akan tampak juga hal-hal apa saja yang perlu diperbaiki dari model yang sudah ada.

3. TINJAUAN PUSTAKA

Sejak diperkenalkan pada 2018, Transformers merupakan arsitektur *deep learning* yang makin banyak digunakan dalam *downstream NLP tasks* seperti klasifikasi dokumen, *neural machine translation* maupun *named entity recognition (NER)* dan *biomedical NER*. Popularitas Transformers ini didasari dari hasil uji coba empiris yang menunjukkan bahwa arsitektur ini mampu mendapatkan nilai tertinggi dalam berbagai data *benchmark* yang dilakukan (Tunstall, et al., 2022). Dalam uji coba pembangunan NER dalam domain umum Bahasa Indonesia, Transformers juga masih mendapatkan nilai tertinggi dibandingkan model *deep learning* yang lain (Santoso, et al., 2021).

Model berbasis Transformers menggunakan pendekatan yang berbeda dibandingkan arsitektur

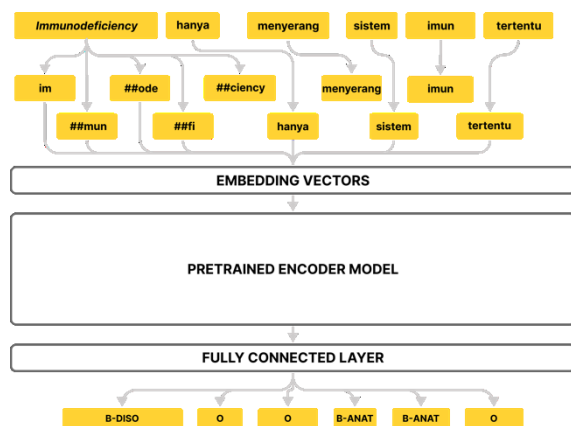
populer lain seperti LSTM. Jika arsitektur LSTM mengirim informasi *hidden states* antar *cell* secara sekuensial, Transformers memanfaatkan konsep *Attention* yang memberikan akses langsung terhadap informasi seluruh *states* tanpa melalui proses sekuens data (Vaswani, et al., 2017). Transformers dapat diimplementasikan ke dalam *encoding* maupun *decoding* task dari data sekuensial seperti teks. Jika diberikan variable Q adalah output vektor dari suatu kata, K adalah seluruh vektor dari kata-kata di dokumen dan V adalah vektor kata sebelum transformasi K , maka nilai *attention* suatu kata dapat dihitung dengan Persamaan 1.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Dimana K^T adalah nilai transformasi dari K , dan $\sqrt{d_k}$ merupakan besaran dimensi vektor *key* dalam (Vaswani, et al., 2017), yaitu 64.

Dengan memanfaatkan prinsip tersebut, seluruh informasi kontekstual dalam suatu korpus dapat tersimpan dengan baik dalam *hidden states*, yang juga disebut *pretrained model*. BERT dan XLM-R merupakan contoh *pretrained encoder model* terdiri dari beberapa *layer* Transformers yang banyak digunakan dalam melakukan *encoding* informasi teks untuk NLP tasks.

Pretrained model merupakan suatu model yang telah dilatih dalam data yang besar untuk menyimpan informasi kontekstual dari seluruh katanya. Dengan memanfaatkan *pretrained model*, pembangunan NLP tasks tidak memerlukan proses training yang lama. Sehingga cukup menambahkan *layer* klasifikasi pada akhir *pretrained model* yang juga dikenal sebagai *fine tuning / transfer learning* (Purwitasari, et al., 2021). Jika diberikan data dalam suatu domain umum D_S dengan *task* NER T_S dan *task* BioNER T_T , maka *fine tuning* akan memaksimalkan fungsi prediktif $f_S(\cdot)$ dalam *target domain* biomedis D_T , dimana $D_S \neq D_T$ dan $T_S \neq T_T$.



Gambar 3. Ilustrasi proses *transfer learning* dari *pretrained model* untuk melatih model BioNER.

Pada konteks pembangunan BioNER, *transfer learning* yang dilakukan adalah menambahkan *fully-connected layer* pada akhir *pretrained model*

sebanyak jumlah label biomedis yang ingin diprediksi seperti yang terlihat pada Gambar 1. Kemudian, data anotasi dari domain biomedis akan dilatih ke dalam model tersebut agar *weights* yang berasal dari model BioNER dapat mengenali label yang telah diberikan. Sedangkan perbedaan antara *pretrained model* seperti BERT dan XLM-R hanya berada pada jumlah *layer* Transformers, jumlah parameter dan optimasi lain yang digunakan model tersebut (Conneau, et al., 2020).

Selain dapat digunakannya *pretrained model* untuk melatih model BioNER lebih cepat, kelebihan lain dari Transformers adalah digunakannya teknik tokenisasi *subword* sebagai representasi teks ke vektor dan mengatasi masalah *Out of Vocabulary* (OOV). Teknik ini akan memecah kata yang tidak dikenal ke dalam beberapa *sub-word* yang dikenal. Sebagai contoh, kata *immunodeficiency* akan dipecah menjadi {*im, mun, ode, fi, ciency*} yang telah dipelajari sebelumnya (Gambar 1). Hasilnya, meski terdapat kata yang belum dipelajari oleh model tersebut, nilai vektor katanya tidak bernilai nol. Dengan kelebihan-kelebihan tersebut, penelitian ini akan mengevaluasi penggunaan Transformers sebagai basis pengembangan BioNER berbahasa Indonesia.

4. HASIL DAN DISKUSI

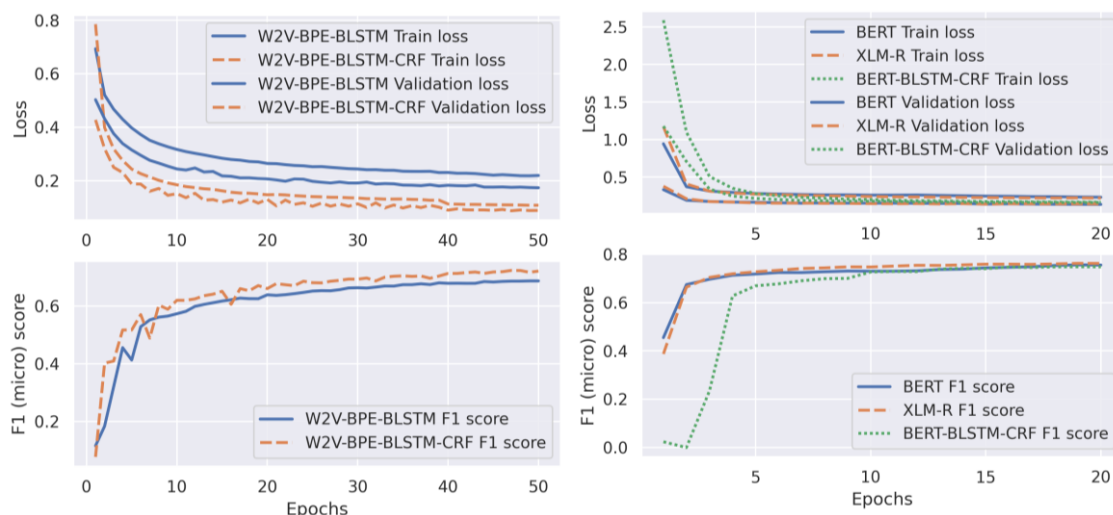
4.1. Representasi Teks Konsultasi Kesehatan Online

Untuk menganalisis representasi teks KKO dibandingkan dengan data media sosial, penelitian ini menggunakan analisis kualitatif yang ditunjukkan pada Tabel 3 berikut.

Tabel 3. Perbandingan antara data teks KKO (atas) dan data Twitter (bawah). Kata-kata yang digarisbawahi merupakan entitas biomedis.

No	Dokumen Data KKO
1	Sudah hampir dua bulan saya <u>sakit kepala sebelah kanan bawah</u> , dok. Saya merasa <u>tegang</u> dan <u>rasa sakitnya menjalar ke bagian atas kepala</u> saya. Ketika saya melihat ke kanan juga <u>terasa sakit</u> . Ada apa yang dok?
2	Halo dok, saya menderita <u>asma</u> dan <u>bronkitis</u> dan tahun ini saya juga mengidap <u>TBC</u> . Diperkirakan saya juga dapat <u>reaksi alergi batuk</u> dan <u>sesak nafas</u> waktu makan, minum yang mengandung susu, daging, makanan laut, dan tepung...
	Data Twitter
1	Mestinya kamu bakal <u>pusing</u> waktu di pesawat nanti ☹️
2	Aku harap gak ada yang dapat <u>covid</u> deh. <u>Badan</u> saya <u>sakit semua</u> , <u>demam terus</u> , <u>lidah mati rasa</u> . Gak bisa makan juga...

Dari hasil di atas, tampak bahwa data dari KKO memiliki konteks bahasan yang lebih jelas dibandingkan dengan data Twitter. Dokumen pertama pada KKO secara eksplisit menunjukkan bahwa pengguna ingin bertanya tentang masalah di kepala. Begitu juga dengan data kedua yang ingin

Gambar 4. Perubahan *Training loss* dan *F1 score* pada LSTM model (kiri) dan Transformers model (kanan)

berkonsultasi tentang penyakit pernafasan. Di sisi lain, dokumen pertama dari Twitter memiliki makna yang ambigu, karena tidak dimaksudkan untuk membahas kesehatan. Tetapi pada dokumen kedua, cuitan lebih sesuai dalam konteks domain kesehatan. Batasan karakter yang hanya 280 karakter juga akan membatasi penyampaian konteks, dibandingkan dengan rata-rata 352 karakter yang dimiliki oleh data KKO.

Secara umum, data dari Twitter membutuhkan tahap interpretasi tambahan untuk mengerti konteks bahasan cuitan tersebut. Jadi, meski mudah didapatkan, data Twitter sebenarnya memiliki satu tahapan proses yang lebih panjang yang membuat representasi data KKO lebih baik dibandingkan data Twitter karena konteks bahasan yang spesifik. Dengan menggunakan data KKO, data *input* yang digunakan pada proses latih BioNER pun juga lebih maksimal.

4.2. Evaluasi pemodelan BioNER

Pada metodologi penelitian disebutkan bahwa untuk memilih model terbaik, *hyperparameters* menemukan setting berikut menghasilkan performa tertinggi: *hidden_size* = 128; *learning_rate* = 0.2; *dropout_rate* = 6.797e-05; *mini_batch_size* = 32; *epoch* = 50. Untuk arsitektur Transformers, digunakan *learning_rate* = 1e-5; *optimizer* = AdamW; *batch_size* = 32; *epoch* = 10. Sedangkan hasil dari masing-masing model dapat dilihat pada Tabel 4.

Terdapat beberapa analisis yang dapat dilakukan berdasar Tabel 4. Pertama, uji coba dua jenis arsitektur pada arsitektur LSTM mengonfirmasi bahwa layer CRF memberikan kenaikan performa cukup signifikan dibanding dengan model LSTM umum. Dengan ditamhakkannya *layer* CRF di akhir model Word2Vec-BPEmb + BiLSTM, terjadi kenaikan nilai Precision, Recall dan F1 score di model tersebut. Di nilai Precision terjadi kenaikan

0.03 poin, sedangkan nilai Recall naik 0.021 poin. Hal ini mengakibatkan nilai F1 score turut naik sebesar 0.023 dibanding model tanpa CRF layer.

Tabel 4. Hasil evaluasi pada tiap arsitektur

Arsitektur	Precision	Recall	F1
Word2Vec-BPEmb + BiLSTM	0.7060	0.6920	0.6989
Word2Vec-BPEmb + BiLSTM + CRF	0.7335	0.7134	0.7228
BERT + Fine tune	0.7814	0.7390	0.7594
XLM-R + Fine tune	<u>0.7818</u>	<u>0.7572</u>	<u>0.7691</u>
BERT + BiLSTM + CRF	0.7616	0.7402	0.7504

Meski demikian, model LSTM masih berada di bawah Transformers secara keseluruhan. Hasil dari *fine tuning* BERT mendapatkan hasil F1 score 0.7594, sedangkan XLM-R mendapat nilai F1 score 0.7691. Jika dilihat secara umum nilai Precision dari Transformers juga unggul dibandingkan LSTM dengan nilai cukup signifikan (5%-8%). Begitu juga pada Recall score pada yang unggul antara 2%-6%. Seluruh hasil ini didapatkan dengan epoch yang lebih kecil dibandingkan LSTM. Fenomena ini diindikasikan karena penggunaan *pretrained model weights* yang telah dilatih pada korpus data yang besar sebelumnya.

Sementara itu, hal menarik terjadi pada *hybrid* BERT + BiLSTM + CRF model. Meski memiliki arsitektur yang lebih kompleks dibanding dua model sebelumnya, hasil evaluasi terhadap model ini tidak lebih baik dibanding *fine tuning* Transformers biasa. Walau memiliki Recall yang lebih baik dibanding BERT, nilai *Precision* model ini masih sedikit dibawah BERT. Bahkan keseluruhan *Precision* dan *Recall* model ini masih dibawah XLM-R. Hal ini mengindikasikan bahwa kompleksitas model tidak berimplikasi positif terhadap performanya. Hasil ini berbeda dengan temuan dari (Dai, et al., 2019) yang

mengklaim bahwa BERT-BiLSTM memiliki performa yang tinggi. Sehingga pada penelitian ini, model *hybrid* BERT-BiLSTM tidak disarankan.

Selanjutnya adalah analisis terhadap hasil *training loss* dari seluruh model. *Loss score* pada model BiLSTM (sebelah kiri) menunjukkan bahwa terjadi konvergensi yang merata antara fase training dan validation sesuai yang diharapkan. Ini menunjukkan bahwa distribusi data yang pada data training dan validation sebelumnya telah merata. Model yang memiliki layer CRF juga menunjukkan *loss* yang lebih rendah. Grafik perubahan F1 score juga menunjukkan bahwa model Word2Vec-BPEmb + BiLSTM + CRF secara konsisten mendapatkan skor yang lebih tinggi di hampir seluruh *epoch*. Ini mengonfirmasi bahwa CRF memiliki efek yang positif pada model LSTM.

Model Transformers, di lain sisi, memiliki pola yang berbeda. *Loss score* pada model ini menunjukkan penurunan yang tajam mulai epoch ke 2, dan mulai melandai setelahnya. Ini terjadi pada arsitektur BERT dan XLM-R tanpa ada selisih *loss* yang signifikan. Fenomena mendukung hipotesis sebelumnya bahwa penurunan yang tajam disebabkan *weights* yang telah terinisialisasi pada *pretrained model* yang membuat *loss* dapat lebih cepat stabil. Pada akhirnya, hal ini yang membuat nilai F1 score pada BERT dan XLM-R juga tidak memiliki perbedaan signifikan dimana XLM-R hanya memiliki 1% keunggulan dibanding BERT.

Terakhir, model *hybrid* BERT + BiLSTM + CRF menghasilkan pola yang cukup berbeda dari Transformers meski menggunakan Transformers sebagai basis *encoder*. Model ini menunjukkan *initial loss* yang jauh lebih tinggi dibandingkan Transformers. Nilai F1 score juga sempat memiliki skor mendekati nol pada dua epoch awal, jauh dibandingkan dengan Transformers yang mendapatkan nilai diantara 0.4 – 0.5. Meskipun demikian, model *hybrid* ini mampu mencapai *loss score* dan memperoleh nilai F1 yang mendekati Transformers dengan fine tuning. Sehingga pada keseluruhan ini, BioNER berbasis XLM-R yang akan digunakan untuk model pengenalan entitas biomedis pada tahap simulasi data.

4.3. Simulasi data dan analisis hasil

Data latih model BioNER diambil dari empat kategori penyakit seperti yang dijelaskan sebelumnya. Jadi, untuk melakukan simulasi data agar menggambarkan performa riil, data simulasi tidak dilakukan filter kategori penyakit untuk melihat generalisasi yang dilakukan model ini. Karena panjangnya dokumen sebagai simulasi, maka terdapat beberapa bagian dari dokumen yang dihapus pada laporan ini. Entitas yang ditandai dalam huruf tebal adalah DISO, sedangkan entitas bergaris bawah adalah ANAT.

Tiga data pertama pada Tabel 5 memberikan gambaran proses ekstraksi pada dokumen KKO.

Hasil ekstraksi menunjukkan bahwa model mampu mengekstraksi entitas ANAT dan DISO pada seluruh dokumen. Temuan menarik dari proses ini adalah model mampu mengenali gejala penyakit yang disampaikan secara informal seperti entitas “seperti ada yang nyangkut” dan “ngringsek gitu kedalam”. Kata-kata tadi merupakan istilah umum yang digunakan oleh masyarakat umum yang seringkali tidak tertangkap *keyword seeds* pada data sosial media. Selain itu, model BioNER yang dibuat juga mampu mengenali entitas biomedis pada kesehatan mental (“kesal”, “benci”, “merusak barang”). Topik tentang kesehatan mental seringkali disampaikan dalam bentuk yang implisit sehingga menjadi tantangan tersendiri dalam studi pengenalan sentimen (Zhang, et al., 2017). Oleh karena itu, model ini juga berpotensi dapat digunakan dalam penelitian terkait analisis sentimen.

Tabel 5. Simulasi ekstraksi informasi pada data KKO, Twitter dan Bahasa Inggris

Dokumen	Hasil ekstraksi
Data KKO	
Dok, kenapa mata saya sering keluar lendir putih/bening? Kadang bola mata bagian atas seperti ada yang nyangkut...	Lendir putih/bening, seperti ada yang nyangkut, cairan
Malam hari dok, suami saya sering sakit kepala sebelah kanan, bisa beberapa kali sehari...	Sakit kepala sebelah kanan; ngingsrek gitu kedalam
Halo dok, saya ingin bertanya tentang diri saya yang sering merasa kesal dan benci...	kesal; benci; merusak barang
Data Twitter	
padahal udh sering muncul tweet kakaknya tntg nunda makan dll tapi selalu diabaikan haha ternyata jengjeng kena batunya 2minggu perut meilit parah sampe sekarang nyeri dan mulut ngerasanya asem trs dan berujung dpt obat juga bahu sampe punggung sebelah kiri nyeri. lengan jg pegel. u know lah deh ini kenapa begini	makan; perut meilit; nyeri; mulut ngerasanya asem
Emak sama kk gua batuk pilek, tapi di rumah yang pake maskernya gua.	bahu sampe punggung sebelah kiri nyeri; lengan jg pegel; batuk; pilek
Data Forum Bahasa Inggris	
I have a hard mass on the right side of my knee. I can see a clear groove on my left knee, but there is mass on my right knee. I hadn't really paid attention to it before today...	hard mass on the right side of my knee; clear groove; my left knee; mass; on my right knee; left knee; fatty lipoma tumours.
i have a slightly deviated septum and a get anxious from the breathing difficulties (anxiety attack symptoms). I wake up in the night from not being to breathe, ...	a slightly deviated septum; get anxious from the breathing difficulties; anxiety attack; worse; allergy; slight trouble swallowing
I really don't want to live anymore. I'm not looking forward to anything. I just want to die. I'm not in danger of hurting myself, I just want to be dead. I'm so tired. So tired of living.	hurting; tired

Selanjutnya, model digunakan dalam uji coba ekstraksi dalam data Twitter. Pada uji coba ini, BioNER yang dibentuk juga berhasil mengekstraksi

entitas biomedis yang dari cuitan pengguna. Pada cuitan pertama, model berhasil melakukan ekstraksi DISO, namun terjadi *entity misclassification* pada kata “makan” dimana itu bukan suatu gejala penyakit. Dikenalnya entitas “lengan jg pegel” pada cuitan kedua juga semakin menunjukkan petunjuk bahwa model tidak terbatas pada bahasa formal. Pada cuitan ketiga, entitas gejala penyakit berhasil dikenali dengan baik.

Pada akhir analisis performa model, penelitian ini juga menguji kemampuan multibahasa karena basis model BioNER menggunakan *pretrained model* XLM-R. Skenario uji coba dilakukan dengan mengambil *sample* diskusi yang ada dalam situs patient.info untuk dikenali entitas biomedisnya.

Hasilnya, model juga dapat melakukan pengenalan entitas biomedis dengan baik. Entitas DISO dan ANAT pada dokumen pertama dapat dikenali dengan baik. Entitas tumor (“*fatty lipoma tumours*”) juga dapat diekstraksi meskipun dokumen tidak dilatih pada data terkait tumor. Proses ekstraksi untuk permasalahan kesehatan mental pada dokumen kedua dan ketiga juga berhasil dilakukan. Namun pada dokumen ketiga, entitas tidak berhasil dilakukan dengan sempurna dengan terlewatkannya beberapa gejala penyakit kesehatan mental seperti “*don’t want to live anymore*” (tidak ingin hidup lagi) dan “*just want to die*” (ingin mati saja). Hal ini dapat dikarenakan makna implisit yang telah disebutkan sebelumnya, serta data latih yang tidak dilatih pada domain kesehatan mental secara khusus.

Evaluasi hasil juga menunjukkan bahwa model masih memiliki beberapa hal yang bisa ditingkatkan. Pertama, pengenalan entitas biomedis yang disampaikan secara implisit dapat dikembangkan. Kedua, terbatasnya entitas pengenalan ANAT dan DISO membuat sistem belum dapat menangkap seluruh informasi biomedis yang terkait pada kalimat, seperti prosedur medis, zat-zat kimia, serta nama penyakit. Ketiga, masih terdapat juga *entity misclassification* entitas biomedis seperti “kehamilan” dan “makan” yang terdeteksi sebagai DISO. Ini membuat pengenalan entitas perlu dioptimasi untuk mengurangi *false positive* klasifikasi.

Secara keseluruhan, hasil analisis kualitatif terhadap ekstraksi entitas menggunakan BioNER dapat dilakukan dengan baik. Model BioNER dapat mengenali entitas biomedis DISO dan ANAT secara general meski hanya dilatih pada empat kategori penyakit. Adanya variasi bahasa informal dalam dokumen simulasi juga tidak mempengaruhi hasil ekstraksi. Penyampaian *disorders* secara naratif seperti “mulut ngerasanya asem” juga dikenali dengan baik. Ditambah juga, karena *pretrained model* yang digunakan mendukung inferensi multibahasa, model juga dapat mengenali entitas biomedis bahasa Inggris. Faktor-faktor tadi membuat BioNER yang dibuat cukup *robust* untuk digunakan dalam berbagai macam skenario ujicoba lain.

5. KESIMPULAN

Ekstraksi informasi merupakan suatu tahapan penting dalam penelitian berbasis teks di domain kesehatan. Oleh karena itu, penelitian ini mengusulkan suatu model pengenalan entitas medis (BioNER) yang dibangun dari data KKO dengan menggunakan arsitektur Transformers. Berbeda dengan penelitian-penelitian sebelumnya yang banyak memanfaatkan data sosial media, penelitian ini menunjukkan bahwa penggunaan data konsultasi kesehatan sebagai data latih BioNER memiliki representasi yang lebih baik. Evaluasi empiris terhadap arsitektur *deep learning* juga menunjukkan bahwa Transformers memiliki performa yang lebih baik dibanding dengan LSTM untuk mengenali entitas biomedis dengan nilai F1 score mencapai 0.7691 untuk model XLM-R. Selain itu model yang lebih kompleks seperti BERT + BiLSTM + CRF tidak memiliki performa yang lebih baik dibandingkan model *fine tuning* standar.

Simulasi juga dilakukan untuk melihat bagaimana BioNER melakukan ekstraksi informasi pada data riil. Hasilnya, beberapa temuan penting dapat diketahui. Pertama, BioNER mampu mengenali entitas biomedis dalam topik yang luas seperti meskipun hanya dilatih pada data konsultasi penyakit HIV, AIDS, diare dan TBC. Kedua, model tersebut juga tidak dipengaruhi oleh variasi bahasa yang disampaikan oleh pengguna. Dengan kemampuan ini, simulasi terhadap data sosial media juga dapat dilakukan dengan baik. Ketiga, dengan terpilihnya BioNER berbasis XLM-R, model ini juga dapat mengenali entitas biomedis dalam bahasa Inggris. Hal ini menarik karena model ini berpotensi dapat digunakan dalam ekstraksi informasi multibahasa.

Penelitian ini juga memiliki keterbatasan, terutama dalam pengenalan tipe entitas biomedis yang lain. Panduan dari dokumen UMLS memiliki ratusan kategori label yang bermanfaat dalam ekstraksi informasi pada domain kesehatan. Sehingga potensi penelitian ke depan, penambahan label data latih dan uji coba terhadap berbagai tipe entitas lain dapat ditambahkan untuk memperkaya hasil ekstraksi yang akan dihasilkan. Selain itu, penelitian lanjutan dapat dilakukan untuk meningkatkan performa model dengan menambahkan proses augmentasi data, serta perbandingan dengan metode *machine learning* diluar metode *deep learning* seperti SVM, *decision tree* maupun CRF. Untuk mendukung upaya penelitian lanjutan, model ini juga dapat diakses pada <https://huggingface.co/abid/indonesia-bioner>.

DAFTAR PUSTAKA

- S. WAKAMIYA, MORITA M., KANO Y., OHKUMA T. dan ARAMAKI E., 2019. Tweet Classification Toward Twitter-Based Disease Surveillance: New Data, Methods, and Evaluations. *Journal of Medical Internet Research*, 21(2).

- ZHANG Y., CHEN K., WENG Y., CHEN Z., ZHANG J. dan HUBBARD R., 2022. An intelligent early warning system of analyzing Twitter data using machine learning on COVID-19 surveillance in the US. *Expert Systems With Applications*, Volume 198.
- BLACK C. M., MENG W., YAO L. dan MILED Z. B., 2022. Inferring the patient's age from implicit age clues in health forum posts. *Journal of Biomedical Informatics*, Volume 125.
- LYU J. C., HAN E. L. dan LULI G. K., 2021. COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *Journal of Medical Internet Research*, 23(6).
- SONG B. , LI F., LIU Y. dan ZENG X., 2021. Deep learning methods for biomedical named entity recognition: a survey and qualitative comparison. *Briefings in Bioinformatics*, 22(6).
- AIELLO A. E., RENSON A. dan ZIVICH P., 2020. Social media- and internet-based disease surveillance for public health. *Annual Review of Public Health*, Volume 41.
- ARACHCHIGE I. A. N., SANDANAPITCHAI P. dan WEERASINGHE R., 2021. Investigating Machine Learning & Natural Language Processing Techniques Applied for Predicting Depression Disorder from Online Support Forums: A Systematic Literature Review. *Information*, Volume 12.
- ATHIRA B., JONES J., IDICULA S. M., KULANTHAIVEL A. dan ZHANG E., 2021. Annotating and detecting topics in social media forum and modelling the annotation to derive directions-a case study. *Journal of Big Data*, Volume 8.
- TUNSTALL L., WERRA L. v. dan WOLF T., 2022. *Natural Language Processing with Transformers*. s.l.:O'Reilly.
- SANTOSO J., SETIAWANA E. I., PURWANTO C. N., YUNIARNO E. M., HARIADI M. dan PURNOMO M. H., 2021. Named entity recognition for extracting concept in ontology building on Indonesian language using end-to-end bidirectional long short term memory. *Expert Systems With Applications*, Volume 176.
- VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER Ł. dan POLOSUKHIN I., 2017. *Attention Is All You Need*. s.l., Advances in Neural Information Processing Systems 30.
- PURWITASARI D., ABDILLAH A. F., JUANITA S. dan PURNOMO M. H., 2021. *Transfer Learning Approaches for Indonesian Biomedical Entity Recognition*. s.l., IEEE.
- CONNEAU A., KHANDELWAL K., GOYAL N., CHAUDHARY V., WENZKE G., GUZMAN F., GRAVE E., OTT M., ZETTLEMOYER L. dan STOYANOV V., 2020. *Unsupervised Cross-lingual Representation Learning at Scale*. s.l., Association for Computational Linguistics.
- CAMPILLOS-LLANOS L., VALVERDE-MATEOS A., CAPLLONCH-CARRIÓN A. Dan MORENO-SANDOVAL A., 2021. A clinical trials corpus annotated with UMLS entities to enhance the access to evidence based medicine. *BMC Medical Informatics and Decision Making* .
- CAMACHO-COLLADOS J. dan PILEHVAR M. T., 2018. *On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis*. s.l., Association for Computational Linguistics.
- HEIJDEN N. V. D., ABNAR S. dan SHUTOVA E., 2020. *A Comparison of Architectures and Pretraining Methods for Contextualized Multilingual Word Embeddings*. s.l., AAAI Conference on Artificial Intelligence.
- JIN G. dan YU Z., 2021. A Korean named entity recognition method using Bi-LSTM-CRF and masked self-attention. *Computer Speech & Language*, Volume 65.
- KOTO F., RAHIMI A., LAU J. H. dan BALDWIN T., 2020. *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*. s.l., International Committee on Computational Linguistics.
- BERGSTRA J., KOMER B., ELIASMITH C., YAMINS D. dan COX D. D., 2015. Hyperopt: a Python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, Volume 8.
- DAI Z., WANG X., NI P., LI Y., LI G. dan BAI X., 2019. *Named Entity Recognition Using BERT BiLSTM CRF for Chinese Electronic Health Records*. s.l., s.n.
- TSAI R. T.-H., WU S.-H., CHOU W.-C., LIN Y.-C., HE D., HSIANG J., SUNG T.-Y. dan HSU W.-L., 2006. Various criteria in the evaluation of biomedical named entity recognition. *BMC Bioinformatics*.
- ZHANG Y., ZHANG O., WU Y., LEE H.-J., XU J., XU H. dan ROBERTS K., 2017. Psychiatric symptom recognition without labeled data using distributional representations of phrases and on-line knowledge. *Journal of Biomedical Informatics*.

Halaman ini sengaja dikosongkan