

Analisis dan Perancangan Filter *Chatbot* untuk Klasifikasi Teks Bermakna Curhat pada Aplikasi Muslim.in menggunakan *Bidirectional Encoder Representations from Transformers* (BERT)

1st Muhammad Izzah Alfatih
AI Center

Fakultas Teknik Elektro
Universitas Telkom

Bandung, Indonesia

izzahalfatih@student.telkomuniversity.ac.id

2nd Suryo Adhi Wibowo
AI Center

Fakultas Teknik Elektro
Universitas Telkom

Bandung, Indonesia

suryoadhiwibowo@telkomuniversity.ac.id

3rd Thomhert Suprpto Siadari
Fakultas Teknik Elektro

Universitas Telkom

Bandung, Indonesia

thomhert@telkomuniversity.ac.id

Abstrak— Individu yang tidak memiliki hubungan sosial yang bermakna berada dalam risiko kesepian, yakni pengalaman rangsangan tidak menyenangkan yang menyebabkan perubahan perilaku. Aplikasi Muslim.in menyelesaikan masalah yang memiliki *chatbot* yang dapat digunakan oleh penggunanya untuk curhat atau menceritakan masalah mereka. Tidak adanya fitur pembatasan topik ketika menggunakan OpenAI API ini menjadikan penggunaan filter sebelum dikirimkan penting. Pembatasan topik obrolan ini berguna untuk memastikan aplikasi digunakan sesuai dengan tujuannya yakni untuk curhat berkaitan dengan permasalahan. *Bidirectional Encoder Representations from Transformers* (BERT) adalah sebuah arsitektur yang berbasis *encoder* dari *transformers*. Arsitektur BERT berhasil meraih *state-of-the-art* baru pada sebelas tugas *natural language processing*. Penggunaan arsitektur BERT pada *dataset* dengan bahasa Indonesia untuk klasifikasi teks juga memiliki performansi yang baik. Pada perancangan ini, model *bert-base-multilingual-uncased* digunakan untuk *fine-tuning* dengan *dataset* berupa kalimat curhat dan bukan curhat. Model awal *bert-base-multilingual-uncased* ditambahkan *dense layer* dengan 256 *neuron* dan *output layer* dengan 1 *neuron*. Model ini dilatih dengan tiga *learning rate* berbeda yaitu 0,01, 0,001, dan 0,0001. Setelah dilakukan pengujian pada *test set*, didapatkan bahwa yang dilatih dengan *learning rate* 0,001 mendapat hasil terbaik. Hal ini dapat disimpulkan bahwa perubahan *learning rate* pada *dataset* berjumlah kecil memiliki pengaruh yang signifikan terhadap performansi model yang dilatih.

Kata kunci— BERT, *chatbot*, kesepian, NLP, *text classification*

I. PENDAHULUAN

Sebagian orang menghadapi kesulitan dalam membangun koneksi sosial yang signifikan, sedangkan sebagian lainnya berhasil melakukannya, tetapi mereka kehilangan hubungan itu karena perpisahan dan banyak hal lain dalam kehidupan [1]. Individu yang tidak memiliki hubungan sosial yang bermakna berada dalam risiko kesepian, yakni pengalaman rangsangan tidak menyenangkan yang menyebabkan perubahan perilaku [1]. Tingkat kesepian memengaruhi gangguan mental emosional secara signifikan [2], menurut penelitian menggunakan analisis regresi pada siswa. Nilai kesepian meningkat 0,274

kali per unit pengukuran, yang berdampak pada gangguan mental emosional [2]. Saat ini, teknologi menawarkan berbagai solusi untuk masalah mental melalui aplikasi *mobile* dan *chatbot* [3], [4]. Studi sebelumnya menunjukkan bahwa *chatbot* juga dapat digunakan sebagai pengganti *Cognitive Behavior Therapy* (CBT) untuk mengurangi depresi dan kecemasan [3]. Selain itu, *chatbot* juga terbukti dapat membantu mengatasi kesepian [5].

Aplikasi Muslim.in menyelesaikan masalah yang memiliki *chatbot* yang dapat digunakan oleh penggunanya untuk curhat atau menceritakan masalah mereka. Penggunaan *chatbot* pada aplikasi Muslim.in dikhususkan untuk curhat berkaitan dengan masalah. *Chatbot* pada aplikasi Muslim.in menggunakan layanan dari OpenAI API untuk membalas pesan dari pengguna. Tidak adanya fitur pembatasan topik ketika menggunakan OpenAI API ini menjadikan penggunaan filter sebelum dikirimkan penting. Pembatasan topik obrolan ini berguna untuk memastikan aplikasi digunakan sesuai dengan tujuannya yakni untuk curhat berkaitan dengan permasalahan.

II. KAJIAN TEORI

A. Text Classification

Text Classification adalah sebuah proses dalam mengkategorikan teks ke dalam grup yang terorganisasi [6]. Tugas yang biasa dilakukan pada klasifikasi teks berupa *sentiment analysis*, *news categorization*, dan *topic classification* [6]. Tujuan dari dilakukannya klasifikasi teks ini adalah untuk mengetahui polaritas atau maksud dari sebuah kalimat. Pada perancangan ini, digunakan model *deep learning* berbasis *transformers* untuk melakukan klasifikasi teks. Penggunaan *deep learning* dengan basis *transformers* juga memungkinkan digunakannya model pra-latih yang sudah dilatih dengan *dataset* yang besar [6]. Performa dari model *deep learning* berbasis *transformers* juga lebih baik dibandingkan dengan model yang lainnya [6], [7].

B. Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers (BERT) adalah sebuah arsitektur yang berbasis

D. Fine-Tuning Model

```

Model: "model"
-----
Layer (type)                Output Shape          Param #          Connected to
-----
Input_ids (InputLayer)      [(None, 256)]         0                (None, 256)
attention_mask (InputLayer) [(None, 256)]         0                (None, 256)
bert (TFBertMainLayer)      TFBertModelOutputLit 167356416        input_ids[0][0]
                                                                attention_mask[0][0]
dense (Dense)                (None, 256)           196864           bert[0][1]
outputs (Dense)              (None, 1)              257              dense[0][0]
-----
Total params: 167,553,537
Trainable params: 197,121
Non-trainable params: 167,356,416
    
```

GAMBAR 5
Susunan Layer Pada Model

Model yang digunakan pada perancangan ini adalah model *pre-trained*. Model *pre-trained* yang digunakan adalah model *bert-base-multilingual-uncased* yang diambil dari *library transformers* yang dibuat oleh *huggingface*. Model diambil dalam format yang sesuai dengan kerangka kerja *tensorflow*. Model dilakukan *fine-tuning* dengan ditambahkan 1 *dense layer* dengan 256 *neuron* dan 1 *output layer* seperti yang terlihat pada Gambar 5. Model dilatih dengan konfigurasi *hyperparameter* seperti pada Tabel 1.

TABEL 1
Konfigurasi Hyperparameter

Hyperparameter	Nilai
Epoch	60
Sequence Length	256
Optimizer	Nadam
Batch Size	16
Learning Rate	0,01 ; 0,001 ; 0,0001

E. Environment Training dan Testing

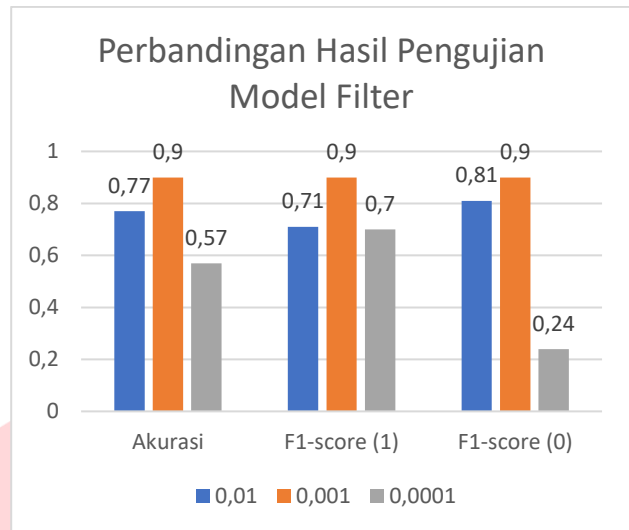
Pelatihan model membutuhkan *library transformers* untuk mengambil *pre-trained bert-base-multilingual-uncased tokenizer* dan model *pre-trained bert-base-multilingual-uncased*. Model ini dijalankan dengan kerangka kerja *tensorflow*. *Library pandas* dan *numpy* juga dibutuhkan untuk mengolah data saat *pre-processing* untuk menggunakan *dataframe* dan *array*. Pada pengujian untuk evaluasi model, dibutuhkan *library* dari *scikit-learn* untuk mendapatkan *classification report* dan *confusion matrix*. Pelatihan dan pengujian dilakukan pada *environment* yang didukung oleh *Compute Unified Device Architecture (CUDA)* pada perangkat komputer personal dengan spesifikasi pada Tabel 2.

TABEL 2
Spesifikasi Perangkat Pelatihan Dan Pengujian

Komponen	Spesifikasi
Operating System	Windows 10 Pro 64-bit
CPU	AMD Ryzen 5 5600G
GPU	NVIDIA GeForce RTX 3060 TI
RAM	16 GB
Storage	512 + 1 TB

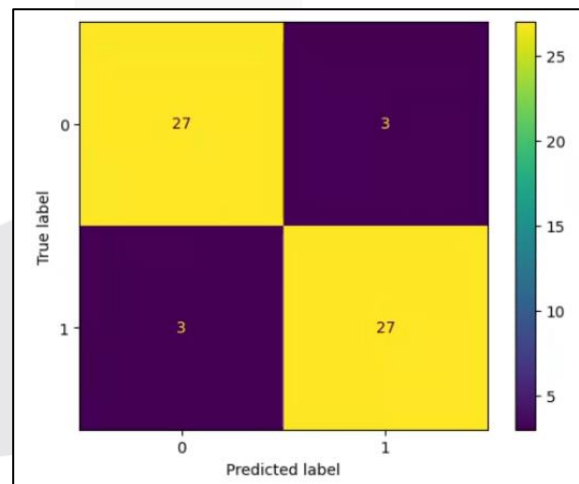
IV. HASIL DAN PEMBAHASAN

A. Hasil Perbandingan Model



GAMBAR 6
Grafik Perbandingan Model

Pada Gambar 6, terlihat grafik perbandingan antara ketiga model. Pengujian nilai akurasi dan *f1-score* dilakukan pada *test set* berisi 60 buah kalimat dengan komposisi 30 kalimat curhat dan 30 kalimat bukan curhat. Terdapat 3 jenis model dengan perbedaan pada *hyperparameter learning rate*. Nilai *learning rate* yang diujikan adalah 0,01, 0,001, dan 0,0001. Berdasarkan hasil pengujian dengan data tes, didapatkan model yang dilatih dengan *learning rate* 0,001 memiliki nilai akurasi dan *f1-score* terbaik dibandingkan yang lainnya. Sehingga, model yang digunakan untuk implementasi adalah model yang dilatih dengan *learning rate* 0,001.



GAMBAR 7
Confusion Matrix

Pada Gambar 7, terlihat kemampuan model terbaik yang dilatih dengan *learning rate* 0,001 dalam memprediksi data sesuai dengan labelnya. Berdasarkan *confusion matrix*, didapatkan nilai TP sebanyak 27 data, TN sebanyak 3 data, *false positive* sebanyak 3 data, dan *false negative* sebanyak 27 data. Hasil dari perhitungan menggunakan fungsi *classification_report* dapat dilihat pada Gambar 8. Model yang dirancang berhasil mendapatkan nilai akurasi sebesar 90%. Nilai dari *f1-score* tiap kelas dapat dilihat juga pada gambar di atas, yakni *f1-score* untuk tiap kelas adalah 0,90.

	precision	recall	f1-score	support
0	0.90	0.90	0.90	30
1	0.90	0.90	0.90	30
accuracy			0.90	60
macro avg	0.90	0.90	0.90	60
weighted avg	0.90	0.90	0.90	60

GAMBAR 8
Classification Report

B. Analisis

Berdasarkan hasil pengujian didapatkan nilai akurasi dan *f1-score* sebesar 0,90 atau 90% dari model dengan *learning rate* 0,001 yang merupakan model dengan performansi terbaik. Nilai ini sudah sesuai dengan yang diharapkan karena hasil keduanya sudah mendekati nilai 1 atau 100%. Dengan kata lain, model yang dirancang telah mampu untuk membedakan mana kalimat yang berupa curhat dan mana kalimat yang bukan curhat. Akan tetapi, pemisahan nilai *float32* hasil prediksi model berada pada nilai 0,3. Hal ini diakibatkan nilai *float32* hasil prediksi mayoritas bernilai 0,3 ke atas. Oleh karena itu, untuk mendapatkan hasil prediksi yang tepat data yang memiliki nilai *float32* hasil prediksi lebih dari 0,3 dikatakan sebagai kalimat curhat.

V. KESIMPULAN

BERT adalah model yang telah mencapai *state-of-the-art* baru pada berbagai tugas NLP. Terlebih lagi, BERT memiliki model yang telah dilatih menggunakan *dataset* dengan berbagai bahasa termasuk bahasa Indonesia pada model *bert-base-multilingual-uncased*. Kemudahan menggunakan BERT untuk berbagai tugas terutama klasifikasi teks dimungkinkan dengan teknik *fine-tuning* pada *pre-trained model* BERT. Pada perancangan ini, model *bert-base-multilingual-uncased* digunakan untuk *fine-tuning* dengan *dataset* berupa kalimat curhat dan bukan curhat. Model awal *bert-base-multilingual-uncased* ditambahkan *dense layer* dengan 256 *neuron* dan *output layer* dengan 1 *neuron*. Model ini dilatih dengan tiga *learning rate* berbeda yaitu 0,01, 0,001, dan 0,0001. Setelah dilakukan pengujian pada *test set* berupa 30 kalimat curhat dan 30 kalimat bukan curhat, didapatkan bahwa yang dilatih dengan *learning rate* 0,001 memiliki hasil yang lebih baik dari dua model lainnya. Hal ini dapat disimpulkan bahwa perubahan *learning rate* pada *dataset* berjumlah kecil memiliki pengaruh yang signifikan terhadap performansi model yang dilatih.

REFERENSI

- [1] C. M. Masi, H. Y. Chen, L. C. Hawkley, dan J. T. Cacioppo, "A meta-analysis of interventions to reduce loneliness," *Personality and Social Psychology Review*, vol. 15, no. 3. SAGE Publications Inc., hlm. 219–266, 2011. doi: 10.1177/1088868310377394.
- [2] C. E. Prasetyo dan T. A. Rahman, "Gangguan Mental Emosional dan Kesepian pada Mahasiswa Baru," *Mediapsi*, vol. 5, no. 2, hlm. 97–107, Des 2019, doi: 10.21776/ub.mps.2019.005.02.4.
- [3] K. K. Fitzpatrick, A. Darcy, dan M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial," *JMIR Ment Health*, vol. 4, no. 2, Apr 2017, doi: 10.2196/mental.7785.
- [4] A. P. Chaves dan M. A. Gerosa, "How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design," *Int J Hum Comput Interact*, vol. 37, no. 8, hlm. 729–758, 2021, doi: 10.1080/10447318.2020.1841438.
- [5] S. Valtolina dan L. Hu, "Charlie: A chatbot to improve the elderly quality of life and to make them more active to fight their sense of loneliness," dalam *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jul 2021. doi: 10.1145/3464385.3464726.
- [6] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, dan J. Gao, "Deep Learning-Based Text Classification," *ACM Computing Surveys*, vol. 54, no. 3. Association for Computing Machinery, 1 Juni 2021. doi: 10.1145/3439726.
- [7] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, dan D. Brown, "Text classification algorithms: A survey," *Information (Switzerland)*, vol. 10, no. 4. MDPI AG, 2019. doi: 10.3390/info10040150.
- [8] J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Okt 2018.
- [9] S. González-Carvajal dan E. C. Garrido-Merchán, "Comparing BERT against traditional machine learning text classification," Mei 2020, doi: 10.47852/bonviewJCCE3202838.
- [10] G. Z. Nabiilah, S. Y. Prasetyo, Z. N. Izdihar, dan A. S. Girsang, "BERT base model for toxic comment analysis on Indonesian social media," *Procedia Comput Sci*, vol. 216, hlm. 714–721, 2023, doi: 10.1016/j.procs.2022.12.188.
- [11] M. N. Fakhruzzaman, S. Z. Jannah, R. A. Ningrum, dan I. Fahmiyah, "Clickbait Headline Detection in Indonesian News Sites using Multilingual Bidirectional Encoder Representations from Transformers (M-BERT)," Feb 2021.