



Membandingkan Nilai Akurasi *BERT* dan *DistilBERT* pada *Dataset Twitter*

Faisal Fajri*, Bambang Tutuko, Sukemi Sukemi

faisal.fajri88@gmail.com*

*Penulis korespondensi

Universitas Negeri Sriwijaya - Indonesia

Diterima: 03 Okt 2022 | Direvisi: 20 Okt – 20 Des 2022
Disetujui: 26 Des 2022 | Dipublikasi: 31 Des 2022
Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Raden Fatah Palembang, Indonesia

ABSTRACT

The growth of digital media has been incredibly fast, which has made consuming information a challenging task. Social media processing aided by Machine Learning has been very helpful in the digital era. Sentiment analysis is a fundamental task in Natural Language Processing (NLP). Based on the increasing number of social media users, the amount of data stored in social media platforms is also growing rapidly. As a result, many researchers are conducting studies that utilize social media data. Opinion mining (OM) or Sentiment Analysis (SA) is one of the methods used to analyze information contained in text from social media. Until now, several other studies have attempted to predict Data Mining (DM) using remarkable data mining techniques. The objective of this research is to compare the accuracy values of BERT and DistilBERT. DistilBERT is a technique derived from BERT that provides speed and maximizes classification. The research findings indicate that the use of DistilBERT method resulted in an accuracy value of 97%, precision of 99%, recall of 99%, and f1-score of 99%, which is higher compared to BERT that yielded an accuracy value of 87%, precision of 91%, recall of 91%, and f1-score of 89%.

Keywords: BERT, DistilBERT, Twitter

ABSTRAK

Ukuran media digital tumbuh dengan kecepatan yang luar biasa, yang membuat konsumsi informasi menjadi tugas yang menantang. Pemrosesan media sosial yang dibantu oleh Machine Learning sangat membantu dalam era yang serba digital. Analisis sentimen adalah tugas mendasar dalam Natural Language Processing (NLP). Berdasarkan banyaknya pengguna media sosial tersebut maka jumlah data yang tersimpan di media sosial juga semakin banyak. Sehingga para peneliti banyak melakukan penelitian dengan memanfaatkan data media sosial ini. Opinion mining (OM) atau Sentiment Analysis (SA) merupakan salah satu metode analisa data media sosial yang mengolah sebuah informasi yang terkandung dalam teks. Sampai saat ini, beberapa penelitian lain mencoba untuk memprediksi Data Mining (DM) menggunakan teknik data mining yang luar biasa. Penelitian ini bertujuan untuk membandingkan nilai akurasi BERT dan DistilBERT. DistilBERT merupakan salah satu teknik dari BERT yang memberikan kecepatan dan memaksimalkan klasifikasi. Dari hasil penelitian, penggunaan metode DistilBERT menghasilkan nilai akurasi sebesar 97%, precision 99%, recall 99% dan f1-score 99%, lebih tinggi bila dibandingkan dengan BERT yang menghasilkan nilai akurasi sebesar 87%, precision 91%, recall 91%, dan f1-score 89%.

Kata Kunci: BERT, DistilBERT, Twitter

PENDAHULUAN

Ukuran media digital tumbuh dengan kecepatan yang luar biasa, yang membuat konsumsi informasi menjadi tugas yang menantang. Sebagian besar media digital dibuat oleh pengguna, tetapi mencari informasi yang diperlukan secara manual berada di luar kemampuan manusia (Gao et al., 2019). Pemrosesan media sosial yang dibantu oleh *machine learning* sangat membantu dalam era yang serba digital. Analisis sentimen merupakan tugas mendasar dalam *Natural Language Processing (NLP)*. *Aspect-Based Sentiment Analysis (ABSA)* disebutkan sebagai tugas terperinci dalam analisis sentimen, yang bertujuan untuk mengidentifikasi polaritas sentimen (misalnya, positif, negatif, netral, konflik) dari kategori aspek atau target (Pontiki et al., 2016).

Perkembangan teknologi saat ini tumbuh sangat pesat, hal ini membuat penyebaran informasi semakin mudah dan cepat melalui media *online* (facebook, twitter), blog, atau situs resmi suatu lembaga (Huddar et al., 2021). Dengan kemudahan dan kecepatan media *online*, mampu mengubah cara konsumsi masyarakat terhadap suatu berita. Berdasarkan banyaknya pengguna media sosial tersebut maka jumlah data yang tersimpan di media sosial juga semakin banyak. Sehingga para peneliti banyak melakukan penelitian dengan memanfaatkan data media sosial ini. *Opinion mining (OM)* atau *Sentiment Analysis (SA)* merupakan salah satu metode analisa data media sosial yang mengolah sebuah informasi yang terkandung dalam teks. *OM* atau *SA* ini merupakan cabang ilmu dari *Text Mining* (Nurrohmat & SN, 2019).

Ada beberapa penelitian yang telah dilakukan sebelumnya dengan menggunakan metode *Bidirectional Encoder Representations from Transformer (BERT)*. Dalam penelitian yang dilakukan oleh Devlin et al. (2019) menggunakan metode *fine-tuning Bidirectional Encoder Representations from Transformer (BERT)* menghasilkan rata-rata nilai akurasi sebesar 82,1%. Geetha dan Karthika Renuka (2021) di dalam penelitiannya masih menghasilkan nilai akurasi yang rendah dibawah 90%. Kemudian, pada penelitian Do dan Phan (2021) nilai akurasi yang dihasilkan sudah di atas 90%. Nilai akurasi tersebut sudah baik, akan tetapi nilai tersebut masih dapat ditingkatkan lagi.

Penelitian ini bertujuan untuk membandingkan nilai akurasi *BERT* dan *DistilBERT*. *DistilBERT* baik digunakan sebagai teknik klasifikasi teks, karena mampu memberikan performa yang sama dengan *BERT* pada berbagai macam tugas pemodelan bahasa alami. Akan tetapi, sebagai tambahan bahwa *DistilBERT* mampu menghasilkan model yang lebih ringan dan lebih cepat pada waktu inferensi (Adoma et al., 2020; Dogra et al., 2021; Preite, 2019; Sanh et al., 2019). Di dalam penelitian ini disajikan hasil pengujian *BERT* dan *DistilBERT* untuk melihat kinerja dari kedua metode. Selain itu, penelitian ini juga ingin membuktikan apakah *DistilBERT* mampu menghasilkan nilai akurasi yang lebih baik dibandingkan dengan *BERT*.

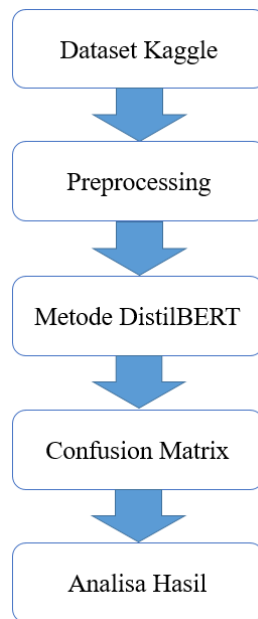
METODOLOGI PENELITIAN

Penelitian ini menggunakan *DistilBERT*, dimana metode ini merupakan pengembangan dari *BERT* yang lebih dulu dikembangkan. *DistilBERT* memiliki arsitektur umum yang sama dengan *BERT* (Ayoub et al., 2021). Data yang digunakan adalah data *twitter* yang mengenai Covid-19, dan nilai yang diukur yaitu akurasi, *recall*, *precision*, dan *f1-score*.

Tahapan Penelitian

Ada beberapa tahapan yang akan dilakukan dalam penelitian ini, dimulai dengan mempersiapkan *dataset twitter* mengenai Covid-19, selanjutnya dilakukan *pre-processing*. Setelah itu dilakukan klasifikasi teks (Joulin et al., 2017), pemrosesan

klasifikasi teks menggunakan *DistilBERT*, dan evaluasi model. Berikut diagram alir tahapan penelitian seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

Penggunaan *Dataset*

Dalam penelitian ini, digunakan *dataset twitter* mengenai Covid-19, *dataset* yang digunakan merupakan *dataset multiclass* yang dimana *dataset* tersebut memiliki sentimen lebih dari 2 kelas, yaitu *positive*, *negative*, *neutral*, *extremely positive*, *extremely negative*. *Dataset twitter* Covid-19 yang digunakan bersumber dari Kaggle. Berikut pada Tabel 1 dan Tabel 2 merupakan penjelasan mengenai *dataset* yang akan digunakan pada penelitian ini.

Tabel 1. *Dataset Twitter*

Sentimen	Jumlah Sentimen	Persentase
<i>Positive</i>	12.343	28%
<i>Negative</i>	10.941	24%
<i>Neutral</i>	8.314	19%
<i>Extremely Positive</i>	7.200	16%
<i>Extremely Negative</i>	6.056	14%

Tabel 2. Pembagian *Dataset*

Jenis Data	Jumlah Data	Persentase
<i>Training</i>	41.106	92%
<i>Testing</i>	3.748	8%

Pada Tabel 1 menjelaskan perbandingan kelas *positive*, *negative*, *neutral*, *extremely positive*, *extremely negative* terhadap data *twitter* mengenai Covid-19. Dari hasil yang

didapatkan dominasi sentimen paling banyak pada kelas *positive* sebanyak 28%, lalu kelas *negative* sebanyak 24%, kelas *neutral* sebanyak 19%, kelas *extremely positive* sebanyak 16%, dan kelas *extremely negative* sebanyak 14%. Kelas *positive* memiliki nilai persentase yang tinggi dibandingkan dengan kelas yang lainnya, ini membuktikan bahwa masyarakat masih memiliki sentimen positif mengenai penyakit Covid-19.

Pada Tabel 2, *dataset* yang digunakan di dalam penelitian ini dibagi menjadi dua bagian, yaitu data *training* sebanyak 41.106 dan data *testing* sebanyak 3.748. Dari *dataset* tersebut digunakan sebagai bahan penelitian untuk menguji nilai akurasi, *precision*, *recall*, dan *f1-score* pada penggunaan *DistilBERT*.

Preprocessing

Sebelum dilakukan pemrosesan data dengan menggunakan *DistilBERT*, dilakukan *preprocessing* terhadap *dataset*. *Dataset* yang didapatkan masih bersifat heterogen dan tidak terstruktur, sehingga diperlukan adanya *preprocessing*. Tujuan dari *preprocessing* adalah untuk mengubah data mentah ke dalam format yang berguna dan efisien. *Preprocessing* data menghadirkan sejumlah tantangan unik yang mengarah ke berbagai algoritma dan teknik heuristik untuk tugas *preprocessing* seperti penggabungan dan pembersihan data, identifikasi pengguna dan sesi.

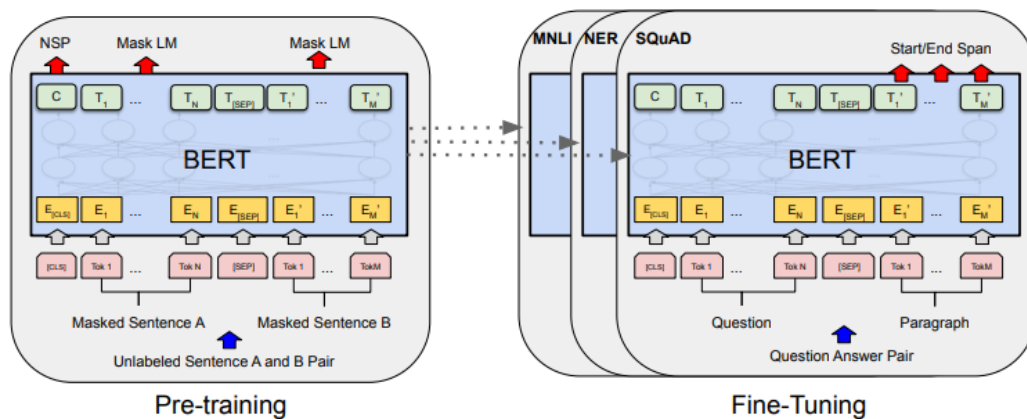
Menurut penelitian yang dilakukan oleh Naseem et al. (2021), *noise* pada kumpulan data *twitter* dapat meningkat hingga 40%, dan ini dapat berdampak signifikan pada kinerja klasifikasi. Oleh karena itu *preprocessing* merupakan salah satu metode untuk menghilangkan *noise* yang terdapat pada kumpulan data *twitter* tersebut. Berikut adalah beberapa teknik yang terdapat pada *preprocessing*, yaitu:

- 1) *Removal noise, URL, Hashtag, dan User Mention*
Merupakan tahapan untuk menghilangkan *hashtag*, *URL*, *mention*, dan tanda baca pada kalimat.
- 2) *Word Segmentation*
Word segmentation adalah proses pemisahan frase atau konten atau kata kunci yang digunakan pada *hashtag*, seperti *#sometrendingtopic* yang disegmentasi menjadi tiga kata yaitu *some+trending+topic*.
- 3) *Replacing Emoticons dan Emojis*
Dalam penelitian yang dilakukan oleh Gimpel et al. (2011), ekspresi dan *emoticon* ini diganti dengan arti kata yang terkait, misalkan :-) diganti dengan senang dan :(dengan sedih.
- 4) *Case Folding*
Case folding adalah tahapan dimana kata yang menggunakan huruf kapital diubah menjadi huruf kecil atau *lowercase*.
- 5) *Tokenization*
Merupakan proses pemisahan setiap kata pada suatu kalimat. Setelah proses tokenisasi maka kata tersebut menjadi sekumpulan array yang setiap selnya berisi kata-kata yang terdapat pada kalimat tersebut (Hermanto et al., 2021).
- 6) *Stemming*
Stemming melakukan proses menghilangkan imbuhan kata pada kalimat sehingga menjadi kata asli (Santosa et al., 2021).

BERT

BERT menggunakan *encoder* dalam *transformator* sebagai sub-struktur untuk model *pre-training* untuk tugas-tugas *NLP* seperti *Sentiment Analysis (SA)*, *Question Answering (QA)*, *Text Summarization (TS)* (Acheampong et al., 2021). Dalam praktiknya, *BERT* melakukan dua fase dalam prosesnya yaitu, *pre-training* untuk pemahaman bahasa

dan *fine-tuning* untuk tugas tertentu. BERT dapat memahami bahasa dengan melatih mekanisme *Masked Language Modeling (MLM)* dan *Next Sentence Prediction (NSP)*. Dua fase yang digunakan di dalam BERT dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur BERT
Sumber: (Devlin et al., 2019)

Selama *pre-training* model dilatih pada data yang tidak berlabel pada tugas *pre-training* yang berbeda. Sedangkan selama *fine-tuning*, model BERT pertama kali diinisialisasi dengan parameter yang telah dilatih sebelumnya, dan semua parameter disetel dengan menggunakan data berlabel dari tugas *downstream*. Setiap tugas *downstream* memiliki model *fine-tuned* terpisah, meskipun diinisialisasi dengan parameter *pre-trained* yang sama.

DistilBERT

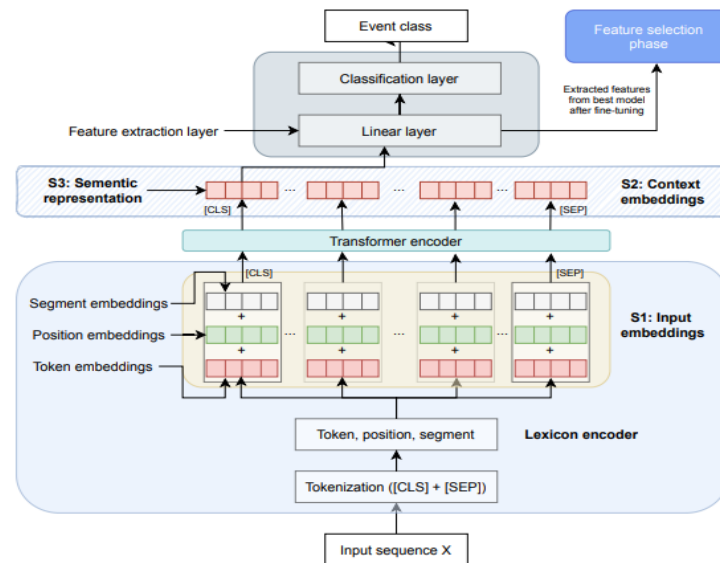
Distillation BERT atau *DistilBERT* dirancang untuk mengurangi ukuran dan meningkatkan kecepatan pelatihan representasi enkoder dua arah dari model *transformer (BERT)*, dimana metode *DistilBERT* menggunakan pengetahuan distilasi (penyulingan) yang bisa meminimalkan model parameter BERT menjadi sebesar 40%, dan membuat inferensi menjadi 60% lebih cepat (Basiri et al., 2021).

Arsitektur model ekstraksi fitur yang menggunakan *DistilBERT* ditunjukkan pada Gambar 3. *DistilBERT* menerima sebagai input X yang mewakili *tweet* dari kumpulan data (urutan kata). Urutan yang dimasukkan ke *DistilBERT* diubah menjadi satu *set* vektor *embedding* dimana setiap vektor dipetakan ke setiap kata secara berurutan (S1).

DistilBERT menggunakan *transformer encoder* untuk mempelajari informasi kontekstual pada setiap kata. *Transformer encoder* menggunakan mekanisme *self-attention* untuk menghasilkan *embedding* kontekstual (S2). *Embedding* kontekstual yang diekstraksi untuk setiap kata digabungkan menjadi satu vektor untuk mewakili informasi semantik yang disajikan dalam *tweet* (S3). S3 merupakan input dari lapisan yang terhubung penuh yang menghasilkan vektor berukuran d di mana d adalah jumlah *neuron*.

Selanjutnya, lapisan klasifikasi ditempatkan di akhir model *ekstraktor* fitur untuk menyempurnakan *DistilBERT* yang telah dilatih sebelumnya pada tugas deteksi *event* dan memprediksi kelas *event* yang sesuai untuk setiap urutan yang dimasukkan (*tweet*). Model *ekstraktor* fitur berada pada lapisan *feature extraction layer* dimana pada layer ini terdapat lapisan linier yang berfungsi untuk mengaktivasi secara linier yang sama dengan input dari lapisan sebelumnya dan keluaran dari fungsi ini adalah inputan + bias. Setelah melewati lapisan linier maka dilakukan tahapan klasifikasi dimana tahapan ini

berfungsi melakukan klasifikasi terhadap setiap inputan yang dihasilkan dari lapisan linier.



Gambar 3. Model Ekstraksi Fitur Menggunakan DistilBERT
Sumber: (Adel et al., 2022)

- 1) *Lexicon Encoder*
Setiap *tweet* diwakili dengan satu *set token* dari vektor *s-length*. *Token* khusus *[CLS]* ditempatkan sebagai *token* urutan (X_1), sedangkan *token* *[SEP]* ditempatkan di akhir urutan. Untuk menghasilkan vektor *embedding* untuk X , *leksikon encoder* meringkas kata, segmen, dan *embedding* posisi untuk setiap token di X .
- 2) *Transformer Encoder*
Menggunakan *DistilBERT*, representasi dipelajari melalui *pre-training*. *DistilBERT* menggunakan distilasi pengetahuan untuk meminimalkan parameter model dasar *BERT* (*bert-base-uncased*) sebesar 40%, membuat inferensi 60% lebih cepat.
- 3) *Fine-Tuning on Event Detection Task*
Dengan asumsi bahwa S_3 merupakan *embedding* kontekstual yang dipelajari oleh token *[CLS]*, yang berfungsi sebagai representasi semantik dari *tweet* input X . Jadi, probabilitas X untuk diklasifikasikan sebagai kelas c (yaitu, *event*) diprediksi sebagai fungsi *Softmax* yang digunakan dalam Persamaan (1).

$$P_r(c|X) = \text{Softmax}(W^T \cdot X) \quad (1)$$

di mana W adalah bobot matriks yang dipelajari selama *fine-tuning* pada model *pre-trained* yang digunakan selama inisialisasi model *ekstraktor* fitur, r adalah jumlah kelas.

- 4) *Feature Extraction Layer*
Keluaran vektor S_3 yang dihasilkan dari lapisan *transformator* terakhir, *DistilBERT*, akan diumpungkan ke lapisan yang terhubung penuh dari ukuran 128 untuk mengurangi dimensi ruang fitur dan kemudian dimasukkan ke lapisan klasifikasi.

Evaluasi Model

Evaluasi model yang akan digunakan untuk mengukur performansi pada metode ini adalah dengan menggunakan *confusion matrix* dimana hasil yang didapatkan berupa

nilai akurasi, *precision*, *recall*, dan *f1-score*. Hasil yang diperoleh dari perhitungan *confusion matrix* akan menjadi acuan apakah penggunaan *DistilBERT* mampu menghasilkan nilai akurasi yang lebih baik jika dibandingkan dengan *BERT*.

1) Akurasi

Akurasi menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar. Akurasi merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (Faturrohman & Rosmala, 2022). Perhitungan akurasi dapat menggunakan rumus pada persamaan (2).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

2) *Precision*

Precision merupakan nilai prediksi antara data yang diminta dengan hasil dari prediksi yang diberikan oleh model. Untuk menghitung nilai *precision* gunakan rumus pada persamaan (3).

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

3) *Recall*

Recall menggambarkan seberapa banyak kelas positif yang diprediksi dengan benar. *Recall* menggunakan rumus pada persamaan (4).

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

4) *F1-Score*

F1-Score merupakan perbandingan rata-rata antara *precision* dan *recall*. *F1-Score* menggunakan rumus pada persamaan (5).

$$F1 - Score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (5)$$

HASIL DAN PEMBAHASAN

Pengujian dengan *BERT*

Pada pengujian ini menggunakan *BERT* dengan menggunakan *dataset twitter* mengenai Covid-19. Hasil pengujian nilai akurasi, *precision*, *recall*, dan *f1-score* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengujian *BERT*

Sentimen	Accuracy	Precision	Recall	F1-Score
<i>Extremely Negative</i>	87%	86%	91%	88%
<i>Negative</i>	87%	86%	85%	86%
<i>Neutral</i>	87%	91%	86%	89%
<i>Positive</i>	87%	85%	86%	86%
<i>Extremely Positive</i>	87%	89%	89%	89%

Dari hasil pengujian pada Tabel 3 dengan menggunakan *BERT*, didapatkan nilai akurasi yang cukup memuaskan untuk klasifikasi teks yaitu 87%. Untuk nilai *precision* yang tertinggi terdapat pada *sentiment neutral* sebesar 91%. Pada *sentiment extremely negative* memiliki nilai *recall* tertinggi sebesar 91%. Untuk nilai tertinggi *f1-score* terdapat pada *sentiment neutral* dan *extremely positive*.

Pengujian dengan *DistilBERT*

Pada pengujian ini, akan melakukan peningkatan akurasi dari *BERT* dengan menggunakan *DistilBERT*. Hasil dari penelitian ini akan dilihat apakah *DistilBERT* mampu meningkatkan nilai akurasi bila dibandingkan dengan menggunakan *BERT*. Pada Tabel 4 ditampilkan nilai dari uji coba dengan menggunakan *DistilBERT*.

Tabel 4. Hasil Pengujian *DistilBERT*

Sentimen	Accuracy	Precision	Recall	F1-Score
<i>Extremely Negative</i>	97%	98%	99%	99%
<i>Negative</i>	97%	97%	98%	98%
<i>Neutral</i>	97%	99%	93%	96%
<i>Positive</i>	97%	97%	98%	97%
<i>Extremely Positive</i>	97%	98%	99%	99%

Dataset yang digunakan sebagai acuan pada pembentukan model klasifikasi *DistilBERT* penelitian ini sama seperti dataset yang digunakan dengan menggunakan *BERT*. Pada Tabel 4 menjelaskan hasil pengujian klasifikasi *DistilBERT*, dimana nilai akurasi yang didapatkan sebesar 97%. Untuk nilai *precision* yang tertinggi terdapat pada sentimen *neutral* sebesar 99%. Untuk *recall* dan *f1-score* nilai tertinggi terdapat pada sentimen *extremely negative* dan *extremely positive*, masing-masing nilai tersebut adalah sebesar 99%.

Dari hasil tersebut, klasifikasi teks dengan menggunakan *DistilBERT* mendapatkan nilai akurasi yang cukup tinggi sebesar 97%. Hasil akurasi yang diperoleh tersebut membuktikan bahwa penggunaan *DistilBERT* dengan menggunakan *dataset twitter* bekerja sangat baik untuk klasifikasi teks.

KESIMPULAN

Berdasarkan hasil pengujian pemrosesan data dengan menggunakan *dataset twitter* mengenai Covid-19 dengan menggunakan *BERT* didapatkan nilai tertinggi untuk akurasi sebesar 87%, *precision* sebesar 91%, *recall* sebesar 91%, dan *f1-score* sebesar 89%. Dengan menggunakan *DistilBERT* pada *dataset* yang sama mampu menghasilkan nilai tertinggi untuk akurasi sebesar 97%, *precision* sebesar 99%, *recall* sebesar 99%, dan *f1-score* sebesar 99%. Dari hasil tersebut diketahui bahwa *DistilBERT* mampu menghasilkan nilai yang tinggi untuk akurasi, *precision*, *recall*, dan *f1-score* pada proses klasifikasi. Dengan hasil yang didapatkan sudah terbukti bahwa *DistilBERT* lebih baik dibandingkan dengan *BERT*.

Perlu penelitian yang lebih mendalam lagi untuk pengembangan klasifikasi teks dengan menggunakan *DistilBERT*, terutama pengurangan waktu pada saat komputasi program. Walaupun waktu komputasi dengan *DistilBERT* sudah mengalami pengurangan tetapi penggunaan *DistilBERT* masih dirasakan terlalu lama pada saat komputasi. Sehingga perlu penelitian lebih lanjut untuk mengurangi proses komputasi program tersebut dengan *DistilBERT*.

DAFTAR RUJUKAN

- Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Recognizing emotions from texts using an ensemble of transformer-based language models. *18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 161–164. <https://doi.org/10.1109/ICCWAMTIP53232.2021.9674102>
- Adel, H., Dahou, A., Mabrouk, A., Elaziz, M. A., Kayed, M., El-Henawy, I. M., Alshathri, S., & Ali, A. A. (2022). Improving Crisis Events Detection Using DistilBERT with Hunger Games Search Algorithm. *Mathematics* 2022, 10(3), 447. <https://doi.org/10.3390/MATH10030447>
- Adoma, A. F., Henry, N. M., & Chen, W. (2020). Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. *17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020*, 117–121. <https://doi.org/10.1109/ICCWAMTIP51612.2020.9317379>
- Ayoub, J., Yang, X. J., & Zhou, F. (2021). Combat covid-19 infodemic using explainable natural language processing models. *Information Processing & Management*, 58(4), 102569. <https://doi.org/10.1016/J.IPM.2021.102569>
- Basiri, M. E., Nemati, S., Abdar, M., Asadi, S., & Acharrya, U. R. (2021). A novel fusion-based deep learning model for sentiment analysis of covid-19 tweets. *Knowledge-Based Systems*, 228, 107242. <https://doi.org/10.1016/J.KNOSYS.2021.107242>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: pre-training of deep bidirectional transformers for language understanding. *Proceedings OfNAACL-HLT 2019*, 4171–4186. <https://aclanthology.org/N19-1423.pdf>
- Do, P., & Phan, T. H. V. (2021). Developing a bert based triple classification model using knowledge graph embedding for question answering system. *Applied Intelligence* 2021 52:1, 52(1), 636–651. <https://doi.org/10.1007/S10489-021-02460-W>
- Dogra, V., Singh, A., Verma, S., Kavita, K., Jhanjhi, N. Z., & Talib, M. N. (2021). Analyzing distilbert for sentiment classification of banking financial news. *Lecture Notes in Networks and Systems*, 248, 501–510. https://doi.org/10.1007/978-981-16-3153-5_53/COVER
- Faturrohman, F., & Rosmala, D. (2022). Analisis sentimen sosial media dengan metode bidirectional gated recurrent unit. *Prosiding Diseminasi FTI*. <https://e-proceeding.itenas.ac.id/index.php/fti/article/view/962>
- Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-dependent sentiment classification with bert. *IEEE Access*, 7, 154290–154299. <https://doi.org/10.1109/ACCESS.2019.2946594>
- Geetha, M. P., & Karthika Renuka, D. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*, 2, 64–69. <https://doi.org/10.1016/J.IJIN.2021.06.005>
- Gimpel, K., Schneider, N., O'connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., & Smith, N. A. (2011). Part-of-speech tagging for twitter: annotation, features, and experiments. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 42–47. <https://aclanthology.org/P11-2008.pdf>
- Hermanto, D. T., Setyanto, A., & Luthfi, E. T. (2021). Algoritma lstm-cnn untuk binary klasifikasi dengan word2vec pada media online. *Creative Information Technology Journal*, 8(1), 64–77. <https://doi.org/10.24076/CITEC.2021V8I1.264>
- Huddar, M. G., Sannakki, S. S., & Rajpurohit, V. S. (2021). Attention-based multimodal contextual fusion for sentiment and emotion classification using bidirectional lstm.

- Multimedia Tools and Applications*, 80(9), 13059–13076. <https://doi.org/10.1007/S11042-020-10285-X/METRICS>
- Joulin, A., Grave, É., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. *The Association for Computational Linguistics*, 2, 427–431. <https://aclanthology.org/E17-2068>
- Naseem, U., Razzak, I., & Eklund, P. W. (2021). A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter. *Multimedia Tools and Applications*, 80(28–29), 35239–35266. <https://doi.org/10.1007/S11042-020-10082-6/METRICS>
- Nurrohmat, M. A., & SN, A. (2019). Sentiment analysis of novel review using long short-term memory method. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(3), 209–218. <https://doi.org/10.22146/IJCCS.41236>
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., & Eryigit, G. (2016). Aspect based sentiment analysis. *SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings*, 19–30. <https://doi.org/10.18653/V1/S16-1002>
- Preite, S. (2019). *Deep question answering: a new teacher for distilbert* [University of Bologna]. <https://amslaurea.unibo.it/20384/1/MasterThesisBologna.pdf>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter*. <https://arxiv.org/abs/1910.01108v4>
- Santosa, R. D. W., Bijaksana, M. A., & Romadhony, A. (2021). Implementasi algoritma long short-term memory (lstm) untuk mendeteksi penggunaan kalimat abusive pada teks bahasa indonesia. *EProceedings of Engineering*, 8(1). <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/14318>