JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika

[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

PERBANDINGAN NILAI AKURASI *DISTILBERT* DAN *BERT* PADA DATASET ANALISIS SENTIMEN LEMBAGA KURSUS

Ade Chandra Saputra [a,1,*], Agus Sehatman Saragih [b,2], Deddy Ronaldo [c,3]
[a,b,c] Universitas Palangka Raya
[1] adechandra@it.upr.ac.id*; [2] assaragih@gmail.com; [3]d.ronaldo@it.upr.ac.id (9pt)
* corresponding author

ARTICLE INFO

Analisis Sentimen, distilBERT, BERT

ABSTRACT

This research aims to implement Sentiment Analysis in Course Reviews using Transfer Learning approach with DistilBERT language model in the context of educational system development. With the rapid growth in the e-learning domain and online course services, user understanding of various courses has become increasingly important for educational institutions. Transfer learning methods, relying on pre-trained N models like DistilBERT, have proven effective in sentiment analysis tasks with good performance and high efficiency.
With the increasing interest in online learning, this study investigates how sentiment approaches can provide deeper insights into course reviews. By applying DistilBERT techniques, it is expected that the system can effectively extract sentiments contained in these reviews, providing comprehensive understanding of users' opinions feelings towards the courses they take.
Through research, it is expected to make a contribution to course providers in enhancing the quality of educational services they offer, providing more detailed and timely feedback to users. The dissemination this research is expected to provide a broader perspective on the application of transfer learning in sentiment analysis, particularly in the context of reviews.

1. Pendahuluan

Dalam membandingkan nilai akurasi antara *DistilBERT* dan *BERT* pada dataset analisis sentimen lembaga kursus, penting untuk memahami perbedaan antara kedua model tersebut serta relevansinya dalam konteks analisis sentimen di lembaga kursus.

BERT (*Bidirectional Encoder Representations from Transformers*) dan *DistilBERT* adalah dua arsitektur transformer yang telah menjadi standar dalam pemrosesan bahasa alami. *BERT* dikenal sebagai model yang sangat besar dan kompleks, sementara *DistilBERT* merupakan versi yang lebih ringkas dari *BERT*, dirancang untuk lebih cepat dan efisien dalam pelatihan serta penggunaan.

Analisis sentimen merupakan proses untuk mengekstrak dan menilai sentimen atau opini dari teks. Dalam konteks lembaga kursus, analisis sentimen berperan penting dalam membantu lembaga untuk memahami perasaan dan umpan balik pelanggan terkait layanan dan kursus yang mereka tawarkan.

Penelitian ini memiliki signifikansi karena memberikan wawasan yang mendalam tentang performa relatif *DistilBERT* dan *BERT* dalam memahami sentimen pelanggan terkait lembaga kursus. Dengan demikian, lembaga kursus dapat menggunakan wawasan ini untuk meningkatkan kualitas layanan mereka dan mengidentifikasi area yang perlu diperbaiki.

Dalam melakukan perbandingan antara *DistilBERT* dan *BERT*, beberapa faktor perlu dipertimbangkan. Pertama, ukuran dan kompleksitas model dapat mempengaruhi waktu pelatihan dan kebutuhan komputasi.

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

BERT memiliki lebih banyak parameter dibandingkan *DistilBERT*, sehingga memerlukan sumber daya yang lebih besar. Namun, *DistilBERT* dapat menjadi pilihan yang lebih efisien jika keterbatasan sumber daya menjadi masalah.

Selain itu, aspek akurasi dan kecepatan juga penting dalam penilaian kedua model ini. Meskipun *BERT* mungkin memiliki performa yang lebih baik karena kompleksitasnya, *DistilBERT* dapat memberikan hasil yang cukup baik dengan waktu yang lebih singkat.

Metrik evaluasi seperti *precision*, *recall*, dan *F1-score* akan digunakan untuk mengukur performa kedua model dalam melakukan analisis sentimen pada dataset lembaga kursus. Metrik-metrik ini memberikan gambaran yang lebih komprehensif tentang keberhasilan model dalam mengklasifikasikan sentimen positif, negatif, atau netral.

Pengelolaan dataset juga menjadi faktor penting dalam penelitian ini. Pembagian dataset dengan baik antara data pelatihan, validasi, dan pengujian merupakan langkah penting dalam memastikan hasil yang dapat dipercaya dan generalisasi yang baik.

Penelitian sebelumnya telah mencoba menggunakan metode Bidirectional Encoder Representations from Transformer (BERT) dalam analisis sentimen. Studi-studi seperti yang dilakukan oleh Devlin dkk (2019) [3] menunjukkan bahwa fine-tuning BERT menghasilkan akurasi rata-rata sebesar 82.1%, sementara penelitian lain seperti yang dilakukan oleh Geetha dan Karthika Renuka (2021) [6] masih memperoleh akurasi di bawah 90%. Namun, penelitian seperti yang dilakukan oleh Do dan Phan (2021) [4] berhasil mencapai tingkat akurasi di atas 90%.

Penelitian ini bertujuan untuk membandingkan kinerja BERT dan DistilBERT dalam analisis sentimen pada dataset lembaga kursus. DistilBERT merupakan alternatif yang menarik karena mampu memberikan hasil yang setara dengan BERT namun dengan model yang lebih ringan dan waktu inferensi yang lebih cepat (Adoma dkk, 2020 [2] ; Dogra dkk., 2021 [4]; Preite, 2019 [7]; Sanh dkk., 2019 [8]). Dalam penelitian ini, kami akan mengevaluasiERT dan DistilBERT untuk melihat performa keduanya, serta menguji apakah DistilBERT mampu memberikan tingkat akurasi yang lebih baik dibandingkan dengan BERT dalam konteks analisis sentimen lembaga kursus.

2. Metodologi Penelitian

Metode yang digunakan dalam penelitian ini menggabungkan analisis sentimen dengan menggunakan teknik transfer learning yang memanfaatkan model DistilBERT untuk mengevaluasi ulasan kursus. Prosedur ini mencakup pengumpulan data, preprocessing data, pelatihan model, dan evaluasi performa untuk mencapai tujuan penelitian.

Pertama-tama, data ulasan kursus dikumpulkan dari sumber-sumber terpercaya seperti platform pembelajaran daring atau situs review terkait. Data harus mencakup berag opini dan sentimen pengguna terap kursus-kursus tertentu. Kemudian, setiap ulasan akan diberi label sentimen (positif, negatif, atau netral) sesuai dengan konteksnya.

Data ulasan akan dilakukan proses preprocessing untuk membersihkan teks dari noise seperti karakter khusus, tanda baca, dan stop words. Selain itu, teks akan diubah menjadi representasi numerik melalui teknik tokenisasi. Langkah ini diperlukan agar model NLP dapat memahami dan menganalisis konten teks dengan lebih baik.

Model DistilBERT, yang merupakan versi ringan dari BERT, akan digunakan dalam transfer learning untuk tugas analisis sentimen. Untuk melakukan hal ini, model DistilBERT akan disesuaikan dengan data

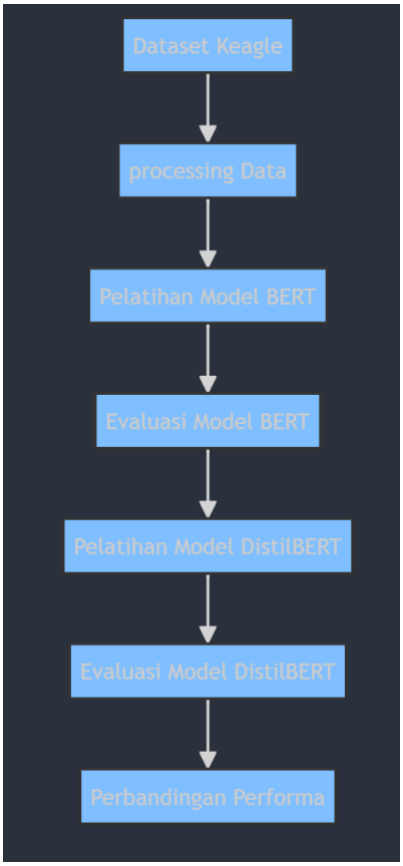
https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

ulasan kursus yang telah diproses sebelumnya. Ini melibatkan masa latih model menggunakan data yang diberi label untuk mengubah representasi kata ke dalam ruang nilai yang menggambarkan sentimen.

Proses analisis sentimen akan memeriksa prediksi yang dihasilkan oleh model DistilBERT terhadap ulasan kursus. Sentimen akan diklasifikasikan menjadi kategori positif,atif, atau net berdasarkan output yang dihasilkan. Metrik evaluasi seperti akurasi, presisi, recall, dan F1-score akan digunakan untuk mengevaluasi performa model.

Data akan dibagi menjadi subset latih, validasi, dan uji. Proses pelatihan dan fine-tuning model akan dilakukan dengan mengoptimalkan parameter model. Eksperimen akan dilakukan untuk menilai keefektifan DistilBERT dalam menganalisis sentimen ulasan kursus, dengan mengevaluasi keakuratan dan generalisasi model. Pada Gambar 1 adalah gambar flowchart atau alur penelitian yang dilakukan.



Gambar 1. Tahapan Penelitian

Preparasi Data

Tahap ini melibatkan pengumpulan data dari sumber Kaggle yang berisi ulasan atau sentimen terkait lembaga kursus. Dataset yang digunakan merupakan dataset multiclass yang memiliki sentimen positif, negatif dan netral. Berikut tabel 1 dan tabel 2 tentang dataset yang digunakan pada penelitian ini.

Tabel 1. Dataset Analisis Sentimen Lembaga Kursus

Sentimen	Jumlah Sentimen
Positive	97227

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

Neutral	5071
Negative	4720

Tabel 2. Pembagian Dataset

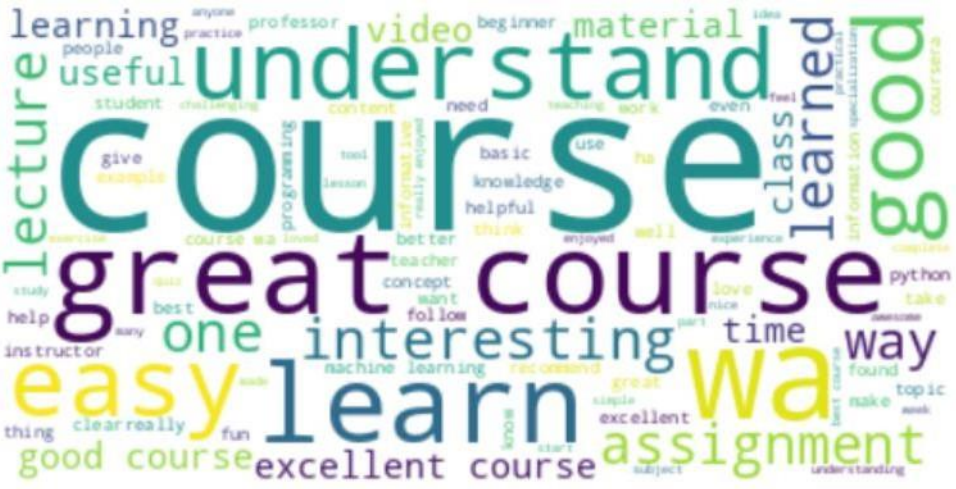
Jenis Data	Jumlah Data
Training	79327
Testing	9916
Validation	9916

Pada Tabel 1 menjelaskan perbandingan kelas positive, negative, dan neutral terhadap data sentimen analisis lembaga kursus. Dari hasil tersebut data sentimen positif mendominasi dengan persentase sebesar 90,85 % , data sentimen negatif sebesar 4,41% dan data sentimen neutral sebesar 4.73%.

Pada Tabel 2 dataset yang digunakan dibagi menjadi tiga bagian yaitu data training sebanyak 79.327, data testing sebanyak 9.916 dan data validation sebanyak 9.916. Dari dataset ini yang akan digunakan untuk menguji nilai akurasi, precision, recall dan f1-score pada penggunaan DistilBERT dan BERT.

Preprocessing data

Sebelum dilakukan pemprosesan data dengan menggunakan DistillBERT dan BERT, dilakukan preprosesing terhadap dataset. Dalam tahapan preprocesing data ini akan dilakukan preprocessing teks agar dapat dilakukan analisis. Tahapan text preprocesing yang dilakukan adalah pemanjangan kata slang inggris, pemanjangan contraction umum pada kata inggris, penghapusan review duplikat, analisis worldcloud dan Checking Sentence Length. Dalam sebuah teks, terdapat banyak kata-kata yang berimbuhan, contohnya adalah walked, walking, dan walks. Sebenarnya, kata-kata tersebut mengandung makna yang sama, yaitu walk (jalan). Oleh karena itu dilakukan lemmatisasi, yaitu teknik untuk mengubah kata yang mempunyai imbuhan menjadi kata dasar. Kemudian, dilakukan penghapusan stopwords agar dapat mendapatkan kata-kata yang bersifat penting dalam analisis. Gambar 2, gambar 3 dan gambar 4 adalah adalah hasil wordcloud analisis yang dilakukan.



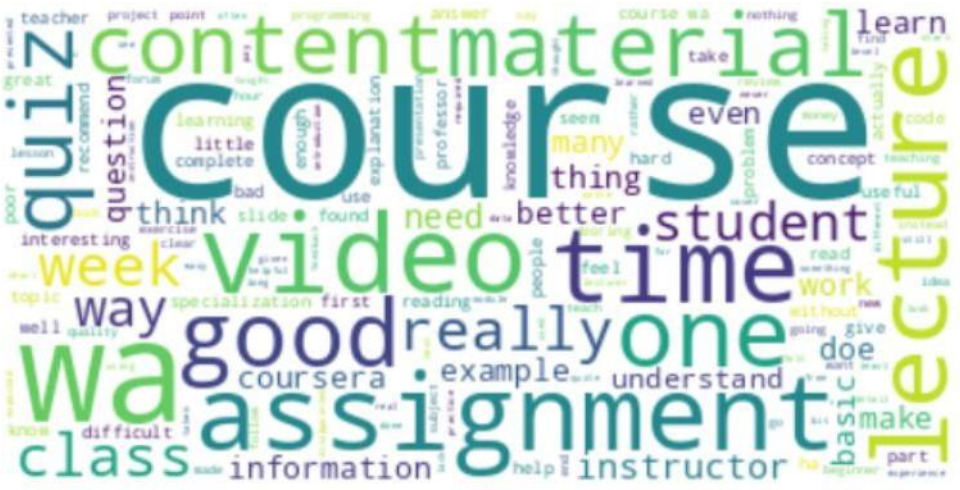
Gambar 2. Wordcloud Analisis Sentimen Positif

Pada Gambar 2 wordcloud analisis sentimen positif, kata yang menjadi sentimen positif adalah course, dan terdapat beberapa kata yang menunjukkan pujian untuk sebuah course di coursera, seperti

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

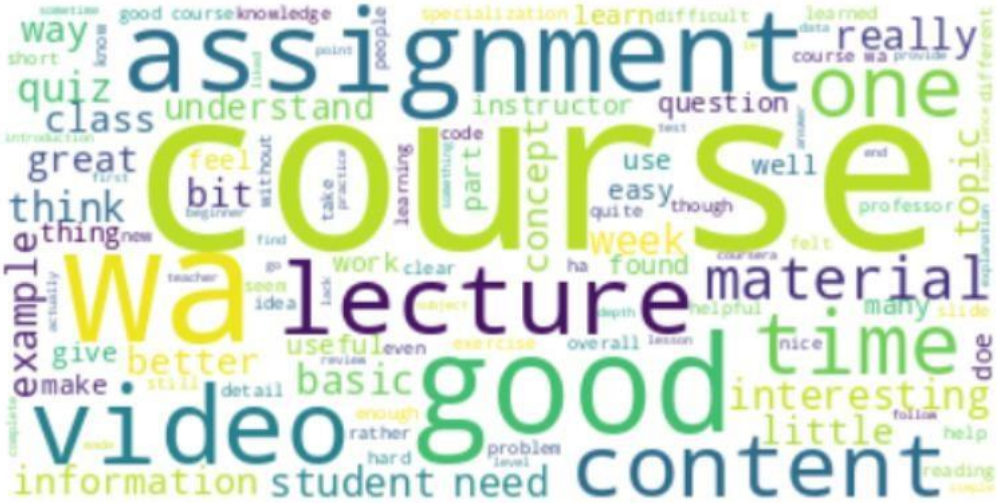
JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

"great course", "understand", dan "excellent course". Kemudian terdapat beberapa aspek yang dibahas meliputi "learning", "wa", dan "assignment". Hal ini menunjukkan bahwa aspek-aspek tersebut merupakan hal-hal yang menjadi kepuasan bagi customer dalam belajar. Hal yang paling menarik dalam wordcloud ini adalah kata "understand", "interesting", dan "easy". Artinya, deliver dari materi yang disampaikan dianggap menarik dan mudah dipahami.



Gambar 3. Worldcloud Analisis Sentimen Negatif

Pada gambar 3 yaitu wordcloud analisis sentimen negatig, terdapat beberapa kata yang muncul adalah kata course, video, assignment, dan lain sebagainya. Ketiga kata ini adalah aspek yang diperhatikan oleh siswa ketika belajar. Kemungkinan besar siswa mempunyai komplain terkait dengan kursus yang dibawakan, assignment yang dikerjakan, serta video materi. Hal yang paling menarik dalam wordcloud ini adalah munculnya kata "understand". Kemungkinan besar siswa punya masalah pemahaman terhadap course-course yang tersedia disana.



Gambar 4. Wordcloud Analisis Sentimen Neutral

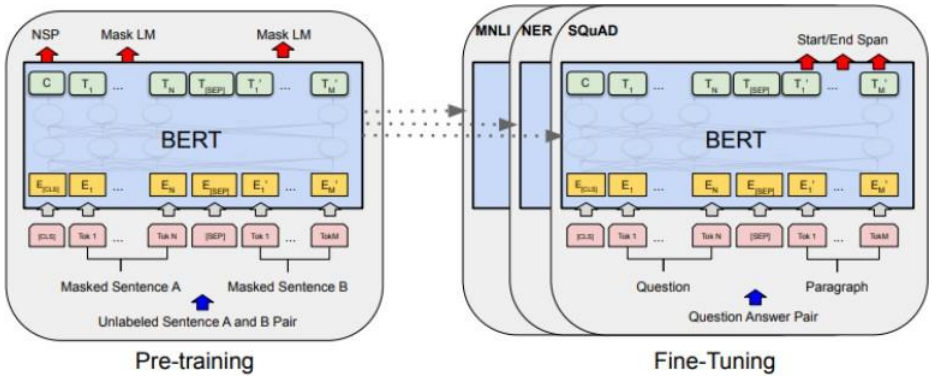
https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

Pada Gambar 4 yaitu worlcloud analsis sentimen neutral terdapat kata "Good", merupakan kata perasaan yang sering muncul. Kemungkinan besar pengguna memuji course tersebut sambil memberikan saran-saran yang membangun.

Pelatihan Model 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 dkk., 2021 [2]). 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 5.



Gambar 5. Arsitektur BERT

Selama pre-training model dilatih pada data yang tidak berlabel pada tugas pretraining 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.

Pelatihan Model 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 6. Seperti pada model NLP lainnya, kita perlu memecah review menjadi kata-kata. Oleh karena itu, kita akan memerlukan tokenizer. Tokenizer untuk membangun model berbasis BERT berbeda dari model lainnya. Proses tokenisasi dapat menggunakan method encode_plus(). Dengan method tersebut, kita dapat menentukan panjang maksimal teks yang akan diproses dan memberikan padding jika ukuran teks kurang dari panjang maksimal. Ketika panjang teks yang asli melebihi maksimal panjang teks, maka teks tersebut akan dipotong. Pada bagian Tokenisasi, diperlukan satu token spesial untuk klasifikasi, yaitu [CLS]. Kemudian kita akan memasukan input id dan attention masknya sebagai masukkan awal model. Ketika sudah masuk ke dalam arsitektur BERT, masing-masing kata akan menghasilkan vektor sendiri-sendiri, termasuk special token [CLS]. Sebagai inputan untuk klasifikasi, kita hanya menggunakan vektor yang

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

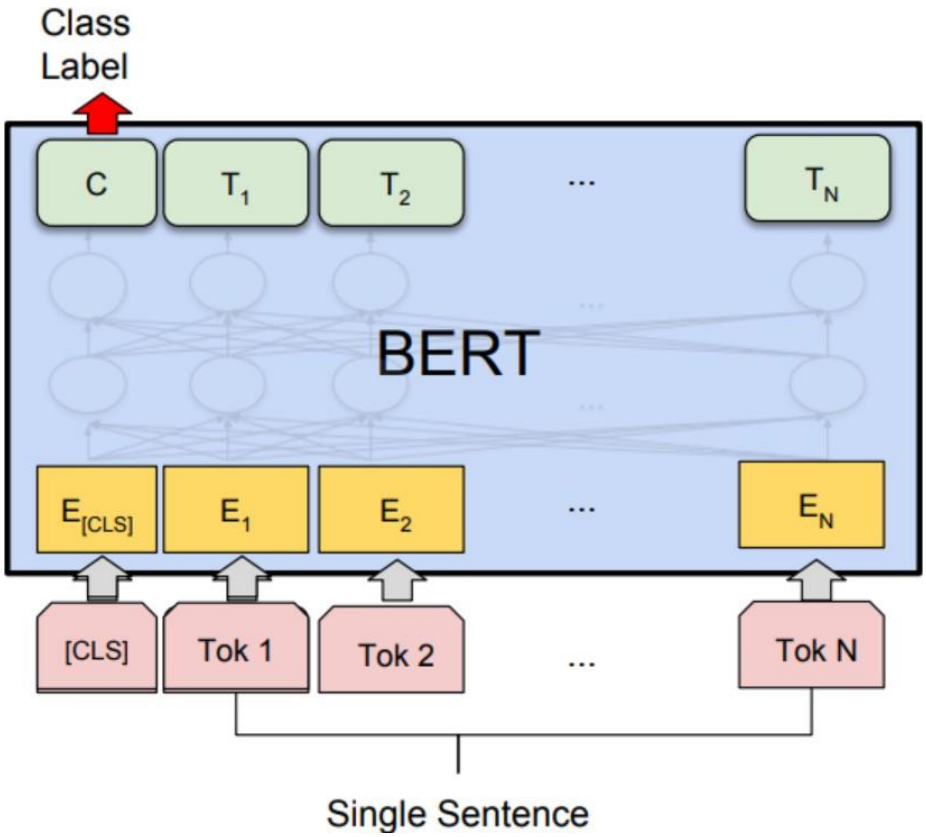
JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

dihasilkan oleh token [CLS] saja. Vektor tersebut mewakili isi dari seluruh teks. Vektor dari token [CLS] adalah inputan untuk layer Fully Connected yang akan digunakan untuk klasifikasi. Pada bagian ini kita menggunakan 2 layer fully connected. Layer pertama mempunyai inputan sebesar 768 dengan outputnya sebesar 512. Kemudian untuk Layer kedua mempunyai inputan sebesar 512 dengan outputnya sebesar jumlah kelas. Karena jumlah kelas adalah 3 maka outputnya adalah 3. Di antara dua layer tersebut terdapat fungsi ReLU.

Dalam model BERT ataupun DistilBERT, terdapat salah satu layer penting, yaitu layer Transformers. Layer Transformers terdiri dari dua bagian, yaitu Encoder (mengekstrak fitur dari teks input) dan Decoder (mengeluarkan output dari fitur yang dihasilkan oleh encoder) .Dalam arsitektur BERT, hanya bagian encodernya saja yang digunakan, sehingga model BERT bisa diartikan sebagai tumpukan dari layer encoder. Untuk inputan awal dari layer encoder, akan ditambahkan salah satu jenis embedding yang bernama Position Embedding, yaitu vektor yang berisi informasi tentang posisi kata. Dalam layer encoder, terdapat salah satu layer yang menjadi layer paling utama dalam arsitektur ini, yaitu multi-head self attention. Dalam arsitektur BERT terdapat 12 layer encoder, sedangkan untuk arsitektur DistilBERT mempunyai 6 layer encoder saja. Hal itu menunjukkan bahwa arsitektur DistilBERT merupakan arsitektur yang lebih ringan bila dibandingkan dengan arsitektur BERT karena mempunyai layer encoder yang lebih sedikit dari arsitektur BERT. Setelah itu, bobot dari model BERT akan di-freeze terlebih dahulu, agar pada saat training bobotnya tidak berubah. Sehingga ketika training, bobot yang akan terupdate adalah 2 layer terakhir saja. Pertimbangan dalam melakukan freeze pada model adalah untuk menghemat waktu komputasi pelatihan model

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]



Gambar 6. Arsitektur Ekstraksi Fitur Menggunakan DistilBERT

Evaluasi Model DistilBERT dan BERT

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 (1).

$$Accuration = \frac{TP+TN}{TP+FP\ +FN+TN} \quad (1)$$

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 (2).

$$Precision = \frac{TP}{TP + H} \quad (2)$$

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

3. Recall
   Recall menggambarkan seberapa banyak kelas positif yang diprediksi dengan benar. Recall menggunakan rumus pada persamaan (3).

$$Recall = \frac{\text{TP}}{TP+FN} \quad (3)$$

4. F1-Score
   F1-Score merupakan perbandingan rata-rata antara precision dan recall. F1-Score menggunakan rumus pada persamaan (4).

$$F1 - Score = \frac{2*(\text{Precision}*\text{Recall})}{\text{Precision}*\text{Recall}} \quad (4)$$

3. Hasil dan Pembahasan

Pengujian dengan DistillBERT

Pada pengujian ini menggunakan BERT dengan menggunakan dataset kaggel mengenai review lembaga kursus. Hasil pengujian nilai akurasi, precision, recall dan f1-score dapat dilihat pada gambar 8. Pada Gambar 7 adalah hasil pelatihan model distilbert.

```
Epoch 1
------------------------------
Train loss 0.8246303053333087 accuracy 0.826981985956862
Val loss 0.715135177681523 accuracy 0.8612343686970553
Done!
Epoch 2
------------------------------
Train loss 0.6745366443485058 accuracy 0.829704892407377
Val loss 0.6780482604618996 accuracy 0.8594191206131504
Done!
Epoch 3
------------------------------
Train loss 0.6512332847698795 accuracy 0.8319109508742294
Val loss 0.66782754015538 accuracy 0.8610326744655102
Done!
Epoch 4
------------------------------
Train loss 0.6404154429985371 accuracy 0.8357557956307435
Val loss 0.6731102836708869 accuracy 0.8689995966115369
Done!
Waktu train DistilBERT: 3875.2219984531403 seconds
```

Gambar 7 . Hasil Pelatihan Model DistillBERT

```
In [46]:
from sklearn.metrics import classification_report

print(classification_report(y_test_lab, y_pred))
```

	precision	recall	f1-score	support
0	0.51	0.72	0.60	460
1	0.22	0.57	0.32	497
2	0.99	0.88	0.93	8959
accuracy			0.86	9916
macro avg	0.57	0.72	0.62	9916
weighted avg	0.93	0.86	0.89	9916

Gambar 8. Hasil pengujian model DistillBERT

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

Model DistilBERT berhasil memprediksi sentimen positif, namun untuk memprediksi sentimen negatif dan netral masih terdapat cukup banyak kesalahan, terutama sentimen netral. Sentimen netral merupakan sentimen yang paling sulit untuk terlihat polanya, karena dalam satu review bisa ada aspek positif dan aspek negatif.

Pengujian Model BERT

Pada pengujian ini, akan melakukan pelatiahan dengan model BERT. Hasil dari penelitian ini akan dilihat apakah BERT mampu meningkatkan nilai akurasi bila dibandingkan dengan menggunakan DistillBERT. Pada Gambar 9 adalah hasil pelatihan dengan model BERT dan gambar 10 adalah hasil pengujian model BERT.

```
Epoch 1
------------------------------
Train loss 0.8134139120674941 accuracy 0.8158382391871619
Val loss 0.6878642431189937 accuracy 0.8328963291649859
Done!
Epoch 2
------------------------------
Train loss 0.6764694848768458 accuracy 0.8329194347447905
Val loss 0.670461000405973 accuracy 0.8489310205728116
Done!
Epoch 3
------------------------------
Train loss 0.6614051324169979 accuracy 0.8358062198242716
Val loss 0.6703568180241892 accuracy 0.8579064138765631
Done!
Epoch 4
------------------------------
Train loss 0.650675627043431 accuracy 0.837949248049214
Val loss 0.6672934837879673 accuracy 0.8659741831383623
Done!
Waktu train BERT: 7200.942486524582 seconds
```

Gambar 9. Hasil Pelatihan Model Bert

```
[66]: print(classification_report(y_test_lab_b, y_pred_b))
```

	precision	recall	f1-score	support
0	0.43	0.75	0.55	460
1	0.21	0.46	0.29	497
2	0.99	0.88	0.93	8959
accuracy			0.86	9916
macro avg	0.54	0.70	0.59	9916
weighted avg	0.92	0.86	0.88	9916

Gambar 10. Hasil Pengujian Model Bert

Setelah mencoba menggunakan model BERT, untuk nilai F1 Score pada ketiga kelas mengalami penurunan sekitar 3%. Namun untuk memprediksi kelas negatif dan kelas netral masih belum baik.

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol 18 No 2]
[Agustus 2024]

Sepertinya perlu penambahan data untuk kelas negatif dan kelas netral agar model dapat mempelajari pola yang lebih beragam.

Hasil Analisis Model

Pada Tabel 3 ini adalah perbandingan antara arsitektur BERT dan arsitektur DistillBERT .

Tabel 3. Perbandingna Model BERT dan Model DistillBERT

Model	Precision	Recall	F1 Score	Accuracy	Waktu Training (s)
Bert	43%	75%	55%	88%	7200.9424
DistillBERT	51%	72%	60%	89%	3875.221

Kedua model tersebut mempunyai akurasi yang mirip. Namun, nilai F1 Score masih cukup rendah. Hal ini disebabkan karena model masih belum bisa menangkap dengan baik kelas minoritas (kelas negatif dan kelas netral) meskipun nilai loss sudah ditambahkan dengan class weight dan kelas netral adalah kelas yang paling sulit diprediksi karena dalam satu teks bisa mengandung makna positif dan makna negatif.

Meskipun seperti itu, waktu eksekusi Arsitektur DistilBERT lebih cepat dibandingkan dengan BERT dan menghasilkan kinerja yang hampir mirip. Sehingga bila perusahaan membutuhkan model yang cerdas namun mempunyai waktu training yang sedikit dapat mencoba menggunakan model DistilBERT. Namun, kita juga perlu melakukan beberapa hal lain untuk meningkatkan performansi model, seperti penambahan review untuk kelas negatif dan kelas netral agar model dapat mempelajari beragam pola yang ada

4. Kesimpulan

Berdasarkan hasil yang didapatkan, kita bisa menarik kesimpulan berikut

1. BERT dan DistilBERT mempunyai nilai F1 Score yang tergolong mirip (hanya berbeda 3% saja). Namun masih belum bisa menangkap kelas minoritas.
2. Menggunakan transfer learning pada model yang "canggih" dapat digunakan untuk task yang mempunyai data berjumlah sedikit. Meskipun sudah mendefinisikan bobot kelas pada loss function namun model belum bisa menangkap kelas minoritas dengan baik. Oleh karena itu kita perlu menambah lagi kelas minoritas atau mengurangi kelas mayoritas supaya model dapat mempelajari kedua kelas dengan baik.

Daftar Pustaka

[1] 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

[2] 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

https://doi.org/10.47111/JTI
Available online at https://e-journal.upr.ac.id/index.php/JTI

Active Media Technology and Information Processing, ICCWAMTIP 2020, 117–121. https://doi.org/10.1109/ICCWAMTIP51612.2020.9317379.

[3] 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.

[4] 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

[5] 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

[6] 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

[7] Preite, S. (2019). Deep question answering: a new teacher for distilbert [University of Bologna]. https://amslaurea.unibo.it/20384/1/MasterThesisBologna.pdf

[8] 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