

Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)

Raden Mas Rizqi Wahyu Panca Kusuma Atmaja¹, Wiyli Yustanti²

^{1,2} Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya

¹radenatmaja16051214009@mhs.unesa.ac.id

²wilyliyustanti@unesa.ac.id

Abstrak— Google Play Store telah menjadi salah satu layanan distribusi digital terbesar yang digunakan untuk mengunduh dan mengunggah aplikasi yang telah di kembangkan. Penelitian ini dilakukan untuk analisa sentimen terhadap aplikasi Ruang Guru di Google Play Store. Data *review* komentar diambil dari fitur komentar yang ada di Google Play Store menggunakan teknik *scrapping*. Data yang digunakan berjumlah 5437 records. Hasil dari analisa sentimen dapat dimanfaatkan untuk melihat respon pengguna Ruang Guru lalu pengembang bisa memaksimalkan fitur yang dirasa kurang oleh pengguna. *Wordcloud* pada komentar review Ruang Guru berisikan keluhan aplikasi, video yang berbayar, dan seringnya aplikasi crash. Model dan metode yang digunakan adalah model *pre-trained* BERT. Pada model ini diperoleh nilai F1 Score adalah 98.9% dengan proporsi data latih dan data uji 70:30. Kemudian, dilakukan evaluasi terhadap model dan menghasilkan nilai akurasi sebesar 99%, presisi sebesar 64.13%, *recall* sebesar 60.51%. Nilai kriteria presentase sentimen memiliki bobot 99% dan dapat dinyatakan valid berdasarkan 10 kali epoch di rata-rata dengan peningkatan yang konsisten.

Kata Kunci— Analisis Sentimen, Ruang Guru, BERT, Klasifikasi, Python, BERT pre-trained model.

I PENDAHULUAN

Ruangguru merupakan perusahaan teknologi terbesar di Indonesia yang berfokus pada layanan berbasis pendidikan. Ruangguru telah memiliki lebih dari 22.000.000 pengguna baik didalam *Appstore* maupun *Playstore* serta mengelola lebih dari 300.000 guru yang menawarkan jasa lebih dari 100 bidang pelajaran. Ruangguru mengembangkan berbagai layanan belajar berbasis teknologi, termasuk layanan kelas virtual, platform ujian *online*, video belajar berlangganan, *marketplace* les privat, serta konten pendidikan lainnya yang dapat diakses baik melalui aplikasi dan web Ruangguru. Ruangguru juga telah dipercaya untuk bermitra dengan 32 Pemerintah Provinsi dan 326 Pemerintah Kota dan Kabupaten di Indonesia. Sehubungan dengan kepopuleran aplikasi Ruangguru beragam reaksi dan tanggapan positif maupun negatif dari pengguna(masyarakat). Tanggapan ini pun dilontarkan baik didalam situs Ruangguru dan *Playstore* atau *Appstore*. Maka dari itu perlu dilakukan analisis *customer review* pada aplikasi Ruangguru untuk mengukur bagaimana respon masyarakat terhadap Ruangguru [1].

Sentiment analysis (analisis sentimen) atau sering disebut juga dengan *opinion mining* (penambangan opini) adalah studi komputasi untuk mengenali dan mengekspresikan opini,

sentimen, evaluasi, sikap, emosi, subjektifitas, penilaian atau pandangan yang terdapat dalam suatu teks [2]. Banyak penelitian diadakan di bidang analisis sentimen karena tingkat persaingan pemasaran yang signifikan ditambah dengan perubahan kebutuhan masyarakat [3]. Penulis memilih *review* Aplikasi Ruang Guru pada Google Play Store sebagai objek penelitian dikarenakan dimasa ini seluruh pembelajaran konvensional sedang ditiadakan dan digantikan dengan pembelajaran daring(online).

Dalam penelitian ini juga menggunakan teori *Data Mining*. *Data Mining* adalah proses ekstraksi informasi dan pola yang berguna dari *Big Data*. Ini juga disebut sebagai proses *knowledge discovery*, penambangan pengetahuan dari data, ekstraksi pengetahuan atau analisis data [4]. Penambahan volume data yang kian pesat dalam digitalisasi, yang berujung konsekuensi tugas seperti pengembangan proses dapat didukung melalui pendekatan *Data Mining* (DM) [5]. CRISP-DM atau Cross-Industry Standard Process for Data Mining merupakan standar yang dikembangkan pada tahun 1996 yang ditujukan untuk proses analisis suatu industry sebagai strategi pemecahan masalah dari bisnis [6].

Penelitian mengenai analisis sentiment telah banyak dilakukan sebelumnya. Beberapa metode telah digunakan diantaranya metode Naive Bayes dalam penelitian [7]. Neural Network dan SVM dalam penelitian [8], SVM dalam penelitian [9]. Penelitian ini akan mengacu pada penelitian [8] tentang pengolahan Customer Review untuk meningkatkan kualitas aplikasi Ruang Guru. Pada penelitian tersebut, Analisa performa dilakukan untuk mengevaluasi kepuasan pelanggan Ruang Guru. Dengan mengekstrak sejumlah *dataset review* pelanggan aplikasi Ruang Guru melalui Google Play Store, kemudian dilakukan *Pre-Processing Text*, Analisis Sentimen, *Top Words Extraction*, *Development of Classification Model* mengacu pada algoritma BERT yang baru saja diciptakan oleh Jacob Devlin pada tahun 2018 [10]. Metode ini sebelumnya sukses dilakukan pada penelitian [11].

Pada penelitian ini terdapat 2 acuan dari segi *Customer review* oleh Alifia [8] dan Algoritma BERT oleh Saad Abdul [11]. Terdapat perbedaan data dan implementasi terhadap subjek yang berbeda. Data yang digunakan oleh penulis menggunakan data *review* komentar Ruang Guru yang di *scrapping* dari Google Play Store, sedangkan penelitian milik Saad Abdul [11] menggunakan data di *IMDB Movie review*. Pada penelitian ini penulis memilih algoritma BERT karena berdasarkan penelitian sebelumnya dinilai memiliki tingkat

akurasi lebih tinggi daripada metode lainnya. Perbedaan penelitian ini dengan penelitian Saad Abdul [11] adalah pada penggunaan *open-source library machine learning*. Pada penelitian ini peneliti menggunakan Pytorch yang dikembangkan oleh Facebook sedangkan untuk penelitian Saad Abdul [11] menggunakan Tensorflow Keras yang dikembangkan oleh Google. Dengan bantuan *wordcloud* dapat membantu peneliti untuk memvisualisasikan hasil data yang tergolong sentimen positif dan sentimen negatif.

BERT adalah deep learning model yang telah memberikan hasil canggih pada berbagai tugas *Natural Language Processing* (NLP). BERT memiliki 6 lapisan Transformer yang ditumpangkan diatas *encoder* dan *decoder* masing-masing, yang menyumbang file proses pelatihan yang sangat kompleks, konfigurasi tinggi, waktu pelatihan yang banyak dan biaya yang sangat mahal. Namun, Google membuka sumber dari pretrained model dari BERT, yang mana dapat penulis gunakan tanpa harus membuat model terlebih dahulu [10]. BERT *processing* dimulai dari sebuah kata dengan representasi *embedding* dari lapisan *embedding*. Setiap lapisan melakukan beberapa penghitungan *multi-headed attention* pada representasi kata dari lapisan sebelumnya untuk membuat representasi perantara baru. Semua representasi perantara ini memiliki ukuran yang sama. Dalam model BERT 12 *layer*, sebuah token akan memiliki 12 representasi menengah [10].

II METODOLOGI

Pada penelitian ini menggunakan algoritma BERT melalui pendekatan metodologi CRISP-DM yang merupakan proses standar terbuka lintas industri untuk penambangan data, metode ini paling banyak digunakan karena penerapannya yang cukup efektif dan memiliki langkah-langkah yang *applicable* (mudah diterapkan).

Urutan langkah-langkah metodologi CRISP-DM akan dijelaskan melalui bagan sebagai berikut:

A. Pemahaman Bisnis & Pemahaman Data

Pada tahap pertama terdapat *business understanding* (pemahaman bisnis) dan *data understanding* (pemahaman data). Pemahaman bisnis membahas terkait tujuan dilakukannya penelitian pada Ruang Guru, disamping itu pemahaman data membahas terkait sumber data dan jumlah data yang akan dimasukkan ke dalam model untuk tujuan penelitian sentimen analisis *review* komentar aplikasi Ruang guru menggunakan metode BERT.

1. Pemahaman Bisnis:

Ruang Guru adalah perusahaan teknologi yang bergerak di bidang Pendidikan NonFormal terbesar di Indonesia yang memiliki lebih dari 22.000.000 pengguna serta telah mengelola lebih dari 300.000 guru yang telah menawarkan lebih dari 100 bidang pelajaran. Fokus peneliti terletak di analisa sentimen *review* komentar Ruang Guru pada platform Google Play Store. Analisa sentimen dilakukan agar dapat mengidentifikasi dan mengklasifikasi komentar mana yang dianggap positif, negatif ataupun netral.

2. Pemahaman Data

Tahap selanjutnya adalah Pemahaman Data. Hal ini penting dilakukan agar data yang akan digunakan dalam penelitian sentimen analisis *review* komentar pada aplikasi Ruang Guru di platform Google Play Store. Tahap pemahaman data sangat berpengaruh terhadap hasil luaran sentimen analisis. Pada proses *machine learning* sentimen analisis seringkali dijumpai data yang tidak konsisten, memiliki *nan-values*, dan masih banyak anomali anomali yang dapat mengganggu proses pelatihan model sentimen analisis. Terdapat beberapa aspek pemahaman terkait pemahaman data yang telah dijelaskan pada alur diatas. Aspek pertama yakni:

a. Sumber Data:

Pada penelitian ini Data yang akan di *scraping*(diambil) berasal dari platform Google Play Store. Dimana Platform ini sudah dipercaya bagi pengguna smartphone android untuk mengunduh dan mengunggah aplikasi.

b. Jumlah data:

Jumlah data yang akan di *scraping* kurang lebih sekitar 12000 *records* yang terdiri dari rating 1-5 pada platform Google Play Store. Peneliti mengambil data dengan jumlah besar karena pada tahap persiapan data akan banyak data yang akan dihapus karena dianggap terlalu banyak mengandung karakter yang tidak bisa dimasukkan ke dalam model, tidak dapat dilakukan pembersihan(*cleaning*).

B. Persiapan Data(Data Preparation)

Sebelum data dimasukkan ke dalam model, perlu dilakukan persiapan data. Hal ini bertujuan agar data yang akan di masukkan ke dalam model sesuai dengan algoritma yang digunakan dalam model tersebut. Persiapan data memiliki beberapa tahap. Dimulai dari proses *scraping* data, proses *cleaning* data, proses labelisasi data, proses pemisahan data menjadi dua bagian latih dan uji, dan proses tokenisasi. Berikut penjelasan tiap tahap dalam persiapan Data:

1. Tahap *scraping* data:

Tahap awal dari persiapan data adalah tahap *scraping* data. Tahap ini menggunakan library *google_play_scraper* sebagai tools untuk mengambil data dan proses ekstraksi

2. Tahap proses *cleaning* data

Proses *cleaning* data kali ini meliputi menghilangkan atribut yang tidak dibutuhkan dalam sentimen analisis seperti ID, waktu komentar, *rating* dll dan atribut lainnya. Berikutnya membersihkan karakter utf-8, titik koma, emoticon, dan proses *text-correction*.

3. Proses *labeling* dan visualisasi data

Proses *labeling* dapat memudahkan model dalam menggolongkan klasifikasi sentimen. Karena input ke dalam model berupa tekstual ke angka jadi data tekstual harus diberi label terlebih dahulu. Proses *labeling* dilakukan secara manual dengan asumsi rating 1-2 tergolong negatif, rating 3 netral, dan

rating 4-5 tergolong positif. Berikutnya dilakukan visualisasi data menggunakan *Wordcloud* untuk melihat tiap kata yang mengandung sentimen tertentu.

4. Proses *splitting* data latih dan uji
Proses ini memisahkan data menjadi dua bagian dengan proporsi 30% : 70%. 30% data uji dan 70% data latih. Pemisahan penting dilakukan karena untuk menentukan data mana yang akan dilatih dan data mana yang akan diujikan sehingga model dapat mengelompokkan berdasarkan data uji dan data latih.
5. Proses tokenisasi dan *data encoding*
Tokenisasi adalah merubah sensitif data menjadi non-sensitif data. Proses ini meliputi: Proses *token embeddings*, *segment embeddings*, dan *position embeddings*. "CLS" adalah token yang dipesan untuk mewakili awal urutan selama "SEP" memisah segment (teks). Proses tersebut adalah:
 - *Token embeddings*: *Token embeddings* secara umum adalah seperti *word embeddings* yakni menggunakan vector untuk mewakili token(kalimat)
 - *Segment embeddings*: *Segment embeddings* memiliki cara kerja ketika masukkan terdiri 2 kalimat, kalimat *embeddings* yang sesuai akan ditentukan ke kata-kata tertentu
 - *Positions embeddings*: *positions embeddings* merferensi kepada urutan token dari *input*(masukkan)Proses ini menjadikan kalimat per kalimat menjadi angka yang kemudian dilakukan proses *encoding* lalu dapat di *input* ke dalam model. Tokenisasi menggunakan library BERT yaitu BERT-Tokenizer.

C. Modeling (Pemodelan)

Pada tahap ketiga ini masuk ke tahap pemodelan, dimana data yang telah dipersiapkan, dibersihkan, dan diberi label akan dimasukkan ke dalam model. Pada penelitian ini model yang digunakan oleh peneliti adalah model *pre-trained BERT classification* dimana model ini telah digunakan oleh penelitian terdahulu dan mendapatkan hasil yang memuaskan dibanding model yang lain.

Tahap pemodelan ini diuraikan menjadi beberapa tahap. Dimulai dengan alur Gbr dan kemudian akan diterangkan secara jelas tiap alur proses pemodelan menggunakan metode BERT

1. Melakukan *set-up BERT pre-trained model*
Pada tahap ini dilakukan instalasi *BERT pre-trained model* sebelum dapat digunakan untuk keperluan sentimen analisis. *Library* yang digunakan adalah *BERTForSequenceClassification*.
2. Melakukan *input* data yang sudah di tokenisasi ke dalam model
Setelah model di *input*, proses selanjutnya yaitu memasukkan data ke dalam model yang sudah di *encode* menjadi token pada proses sebelumnya.

3. Memuat *Data Loaders*
Di dalam proses sentimen analisis yang termasuk kategori *deep learning*, *data loaders* menjadi salah satu komponen penting faktor kesuksesan. *Data loaders* membuat proses pemuatan data lebih efisien untuk skala data ribuan hingga jutaan dikarenakan algoritma yang dimiliki dapat memuat data secara terstruktur dan sistematis dan meminimalisir terjadinya *data loss*.

4. Melakukan *set-up Optimiser and Scheduler*
Dalam melakukan komputasi data, pada umumnya menggunakan *gradient descent* untuk melakukan proses penjumlahan dan turunan dengan tujuan menemukan titik yang meminimumkan suatu fungsi yang dapat diturunkan. Pada penelitian ini menggunakan algoritma *AdamW Optimizer*. Seperti *Stochastic Gradient Descent*, namun algoritma *Adam W* memiliki keunggulan *Root Mean Square Propagation* yang tidak dimiliki di *Stochastic Gradient Descent*. Selanjutnya adalah *Scheduler function*, hal ini memungkinkan untuk menemukan *learning rate* yang optimal pada proses iterasi.

- a) Mendefinisikan Performa Metrik
Pada penelitian ini menggunakan *Scikit-Learn* sebagai software non-berbayar *machine learning library* untuk pemrograman *python*. Pada tahap ini mendefinisikan metrik performa yang menjadi tolok ukur hasil perhitungan komputasi data sentimen analisis. Hal ini meliputi:

- *F1 score*
- *Precision*
- *Recall*
- *Accuracy*

F1 score adalah rata-rata tertimbang dari *precision* dan *recall*. Berikut formula dari *F1 score*:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Precision adalah rasio observasi positif yang diprediksi dengan benar terhadap total observasi positif yang diprediksi. Berikut formula dari *precision* :

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall adalah rasio observasi positif yang diprediksi dengan benar terhadap semua observasi di kelas aktual. Berikut formula dari *Recall*:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Akurasi adalah ukuran kinerja yang paling intuitif dan ini hanyalah rasio observasi yang diprediksi dengan benar terhadap total observasi. Berikut formula dari Akurasi:

$$\text{Akurasi} = \frac{TN + TP}{TP + FN + TN + FN}$$

D. Evaluation(Evaluasi)

Pada tahap keempat terdapat tahap Evaluasi. Tahap ini menjelaskan tentang proses komputasi(*Machine Learning*) terhadap model, data, algoritma,dan fungsi yang telah di definisikan sebelumnya. Namun, proses perhitungan dan sentimen analisis secara rinci didalam python tidak dapat di visualisasikan secara detail. Hanya hasil luaran yang dapat di visualisasikan.

E. Deployment(Peluncuran)

Pada tahap kelima menjelaskan tentang peluncuran. Peluncuran yang dimaksud adalah setelah diketahui hasil parameter dari sentimen analisis, dilakukan dokumentasi(*final project report*). *Final project report* menjelaskan tentang alur dari sentimen analisis mulai pengambilan data, hingga proses komputasi data dan hasil luaran dari komputasi data.

III HASIL DAN PEMBAHASAN

A Data

Data yang dikumpulkan berupa data komentar review sejumlah 10418 records lalu dilakukan *preprocessing* hingga menjadi 5437 records. Teknik pengumpulan data(*scrapping*) dilakukan dengan *library Google-Play-Scrapper*, data yang digunakan diambil dari Google Play Store melalui fitur review komentar. Total data yang berhasil dikumpulkan dapat dilihat pada tabel berikut:

TABEL I
HASIL SCRAPING

Sumber Data	Jumlah	Total
Google Play Store	10418	10418

Untuk rincian kolom data yang telah diperoleh tidak dapat di visualisasikan karena jumlah kolom yang tidak cukup untuk dimuat dalam artikel ini. Data yang telah di *preprocess* dan diberi label disajikan pada tabel 2. Untuk proses pelabelan penulis mengambil berdasarkan rating 1 – 5 dimana rating 1 – 2 di golongan negatif, rating 3 digolongkan netral sedangkan rating 4 – 5 digolongkan positif. Lalu untuk proses sentimen analisis nanti akan di komputasi menggunakan metode BERT dan formula dari *Scikit-Learn* untuk evaluasi label berdasarkan teks apakah teks mengandung sentimen benar sesuai label yang telah diberikan di awal sebelum proses komputasi.

TABEL III
ISI HASIL SCRAPING

ID	Teks	Kategori
Gp12258	Ruang guru is the Best!!!	Positif
Gp00002	"Wow this application so amazing,they help me to get something more than in my school,thank you RUANG GURU"	Positif

ID	Teks	Kategori
Gp00003	bad too much distraction and requirements	Negative

Seluruh data disimpan dalam sebuah file berformat .csv, kemudian dilakukan proses *cleaning* untuk membersihkan atribut, karakter yang tidak dibutuhkan, dan mengkoreksi *misspelling* menggunakan pada *dataset*. Data yang sudah dibersihkan dapat dilihat pada tabel dibawah ini:

TABEL IIIII
HASIL CLEANING

No.	Sebelum	Sesudah
Gp12258	"RUANGGURU IS THE BEST"	ruang guru is the best
GP00009	"Wow this application so amazing,they help me to get something more than in my school,thank you RUANG GURU"	wow this application so amazing they help me to get something more than in my school thank you ruang guru

Setelah proses *cleaning*, data membutuhkan index yang digunakan untuk membaca seluruh elemen yang terdapat pada tabel. Pada tabel berikut terdapat data yang berhasil diberi label:

TABEL IVV
HASIL LABELING

No.	Kategori	Jumlah	Total
1	Positif	5254	5437
2	Negatif	167	
3	Netral	16	

TABEL V
HASIL PENGOLAHAN DATA LABEL

ID	Teks	Kategori	Label
gp00001	a very bad learning application why my ruangguru always loading and it said that they are preparing material for me i had been waiting for more than thirty minutes please fix it	Negative	0
gp00009	a great app for study	Positif	1
gp00402	bribe happen farhan	Netral	2

Berikut juga disajikan bentuk pengolahan data manual terkait dengan penggolongan positif dan negatif. Pengolahan data manual dilakukan didalam *software* Microsoft Excel

2019. Bentuk pengolahan data manual disajikan pada tabel berikut:

TABEL VI
BENTUK PENGOLAHAN DATA MANUAL

ID	Teks	Kategori
gp00001	a very bad learning application why my ruangguru always loading and it said that they are preparing material for me i had been waiting for more than thirty minutes please fix it	negative
gp00002	a good app	positive
gp00003	a good app to learn	positive
gp00004	a good app with lots of good teachers	positive
gp00005	a good application for study easy to used and a lot of advantages to increase the knowledge	positive
gp00006	a good application for study easy to used and a lot of advantages to increase the knowledge	positive
gp00007	a good place to learn boost your knowledge anytime anywhere so easy to use and helpful	positive
gp00008	a good place to learn boost your knowledge anytime anywhere so easy to use and helpful	positive
gp00009	a great app for study	positive
gp00010	a great chance to make yourself different	positive

Berdasarkan isi tabel III, terdapat kategori positif sejumlah 5254 records, negatif sejumlah 167 records, dan netral sejumlah 16 records. Pada proses selanjutnya akan dilakukan proses *splitting* untuk memisahkan data yang akan di latih dan akan di uji. Selanjutnya akan dilakukan proses visualisasi data menggunakan *Wordcloud*, pada proses ini mengklasifikasikan data positif dan data negatif pada Gbr 7.

a) *Wordcloud* Positif



Gbr. 5 *Wordcloud* positif Ruang Guru

Berikut adalah *wordcloud* positif dari dataset ruang guru. Pada *wordcloud* diatas terdapat beberapa kata menonjol seperti “application”, “best”, “good”, “great”. Kalimat tersebut mengGbrkan suka, aplikasi yang bagus dan beberapa pujian lain untuk aplikasi Ruang Guru.

b) *Wordcloud* negatif



Gbr.6 *Wordcloud* negatif Ruang Guru

Berikut adalah *wordcloud* negatif dari *dataset* ruang guru. Pada *wordcloud* diatas terdapat kata menonjol “bad”, “ruangguru”, “paid”, “video”. Kalimat tersebut mengGbrkan bahwa aplikasi buruk, berbayar, dan terdapat beberapa kata “cant access” tidak bisa akses atau sering kali error.

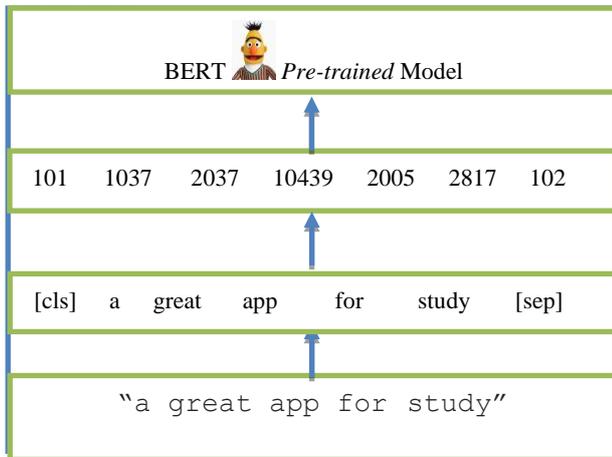
Berikut tabel data yang akan di latih dan diuji dengan proporsi 70:30 dimana 70% data yang akan dilatih dan 30% data yang akan di uji:

TABEL VII
HASIL SPLIT ANTARA DATA LATIH DAN DATA UJI

No.	Kategori	Tipe data	Jumlah
1.	Positif	latih	3677
		Uji	1577
2.	Negative	latih	117
		Uji	50
3.	Netral	latih	11
		Uji	5

Berdasarkan data diatas diperoleh label positif sebanyak 3677 untuk data latih, 1577 untuk data uji. Kemudian untuk

data negatif 117 untuk data latih, 50 untuk data uji. Kemudian untuk data netral terdapat 11 untuk data latih, 5 untuk data uji. Setelah proses pemisahan antara data latih dengan data uji, dilakukan proses tokenisasi data dan dilanjutkan proses encoding data. Berikut alur dan hasil datayang telah di encode ke dalam token:



Gbr.7 Proses input kata-kata menjadi token lalu di input ke dalam model BERT

Pada Gbr 7 dijelaskan alur sebuah contoh kalimat yang dimasukkan menjadi token lalu di input ke dalam model BERT.

B Modeling

Pada tahap ini data telah melalui proses encoding dan tokenisasi. Selanjutnya melalui proses pembuatan model BERT pre-trained yang bertujuan memasukkan dataset ke dalam model untuk di uji dan dilatih. Pada Gbr 8 merupakan hasil set-up BERT pre-trained model.

```

The BERT model has 281 different named parameters.
==== Embedding Layer ====
bert.embeddings.word_embeddings.weight          (30522, 768)
bert.embeddings.position_embeddings.weight      (512, 768)
bert.embeddings.token_type_embeddings.weight    (2, 768)
bert.embeddings.layer_norm.weight              (768,)
bert.embeddings.layer_norm.bias                 (768,)

==== First Transformer ====
bert.encoder.layer.0.attention.self.query.weight (768, 768)
bert.encoder.layer.0.attention.self.query.bias (768,)
bert.encoder.layer.0.attention.self.key.weight (768, 768)
bert.encoder.layer.0.attention.self.key.bias   (768,)
bert.encoder.layer.0.attention.self.value.weight (768, 768)
bert.encoder.layer.0.attention.self.value.bias (768,)
bert.encoder.layer.0.attention.output.dense.weight (768, 768)
bert.encoder.layer.0.attention.output.dense.bias (768,)
bert.encoder.layer.0.attention.output.LayerNorm.weight (768,)
bert.encoder.layer.0.attention.output.LayerNorm.bias (768,)
bert.encoder.layer.0.intermediate.dense.weight (3072, 768)
bert.encoder.layer.0.intermediate.dense.bias   (3072,)
bert.encoder.layer.0.output.dense.weight       (768, 3072)
bert.encoder.layer.0.output.dense.bias         (768,)
bert.encoder.layer.0.output.LayerNorm.weight  (768,)
bert.encoder.layer.0.output.LayerNorm.bias    (768,)

==== Output Layer ====
bert.pooler.dense.weight                       (768, 768)
bert.pooler.dense.bias                         (768,)
classifier.weight                              (3, 768)
classifier.bias                                (3,)
    
```

Gbr.8 Hasil set-up BERT pre-trained model

Setelah model BERT terinstalasi dengan benar, maka dilanjutkan untuk komputasi data ke dalam model dengan beberapa parameter sebagai berikut:

- Batch size : 32 (diatur ketika membuat data loaders)
- Learning rate(Adam): 1e-5
- Epsilon: 1e-5

- Epochs: 10

Pada tahap pemodelan ini, penulis menggunakan Batch size: 32 merujuk pada saran author BERT[10]. Untuk Epoch ditentukan 10 epoch dikarenakan untuk menguji dan memastikan tidak terjadi underfitting (suatu kondisi ketika model tidak mampu untuk mempelajari beberapa data latih yang telah di masukkan) dan akan berakhir dengan performa yang buruk. Terdapat beberapa parameter untuk menentukan apakah model underfitting atau overfitting(suatu kondisi dimana model melewati hasil yang diharapkan pada beberapa iterasi pertama yang dinilai terlalu optimis). Penentuan semua parameter diatas pada dasarnya berdasarkan “Hit-and-trial error” hingga mendapatkan performa yang diharapkan[12]. Diantara semua parameter, Epoch adalah role yang signifikan terkait dengan performa function[12]. Oleh karena itu penulis menentukan epoch 10 karena dinilai mulai dengan epoch ke 6 hingga epoch ke 10, tidak ada perubahan nilai yang signifikan dari parameter yang telah dimasukkan kedalam model.

1. Use Case 90:10

Use case 90:10 yakni dimana data latih di input 90% dan data uji di input 10%. Berikut hasilnya:

TABEL VIII
HASIL EVALUASI PADA BERT PRE-TRAINED MODEL DENGAN PROPORSI DATA 90:10

Epoch	Akurasi	F1	Presisi	Recall
1	98.1%	97.7%	66.04%	49.01%
2	98.7%	98.62%	60.53%	60.59%
3	99.2%	99.1%	66.41%	60.78%
4	98.89%	98.79%	62.24%	60.65%
5	99.08%	98.96%	64.19%	60.72%
6	99.26%	99.16%	64.39%	62.68%
7	99.26%	99.16%	64.39%	62.68%
8	99.26%	99.16%	64.39%	62.68%
9	99.26%	99.16%	64.39%	62.68%
10	99.26%	99.16%	64.39%	62.68%
Average	98.94%	98.89%	64.13%	60.51%

2. Use Case 50:50

Use case 50:50 yakni dimana data latih di input 50% dan data uji di input 50%. Berikut hasilnya:

TABEL IX
HASIL EVALUASI PADA BERT PRE-TRAINED MODEL DENGAN PROPORSI DATA 50:50

Epoch	Akurasi	F1	Presisi	Recall
1	97.79%	97.22%	65.92%	46%
2	97.94%	97.49%	64.32%	48.38%
3	98.45%	98.25%	61.11%	57%

Epoch	Akurasi	F1	Presisi	Recall
4	98.63%	98.48%	96.19%	61.61%
5	98.52%	98.29%	98.88%	58.1%
6	98.63%	98.46%	96.99%	60.84%
7	98.67%	98.48%	98.44%	60%
8	98.48%	98.48%	98.44%	60%
9	98.67%	98.48%	98.44%	60%
10	98.71%	98.52%	98.99%	60.1%
Average	98.44%	98.21%	87.77%	57.2%

3. Use Case 70:30

Use case 70:30 yakni dimana data latih di input 70% dan data uji di input 30%. Berikut hasilnya:

TABEL X
 HASIL EVALUASI PADA BERT PRE-TRAINED MODEL DENGAN PROPORSI DATA 70:30

Epoch	Akurasi	F1	Presisi	Recall
1	98.16%	97.78%	66%	49%
2	98.71%	98.62%	60.53%	60.59%
3	99.26%	99.14%	66.41%	60.78%
4	98.89%	98.79%	62.24%	60.65%
5	99%	98.96%	64.19%	60.72%
6	99.26%	99.16%	64.39%	62.68%
7	99.26%	99.16%	64.39%	62.68%
8	99.26%	99.16%	64.39%	62.68%
9	99.26%	99.16%	64.39%	62.68%
10	99.26%	99.16%	64.39%	62.68%
Average	99%	98.9%	64.13%	60.51%

Pada ketiga percobaan use case diatas penulis mendapatkan rata-rata dari ketiga use case:

Use case 1:

- Akurasi : 98.94 %
- F1 score : 98.89 %
- Presisi : 64.13 %
- Recall : 60.51 %

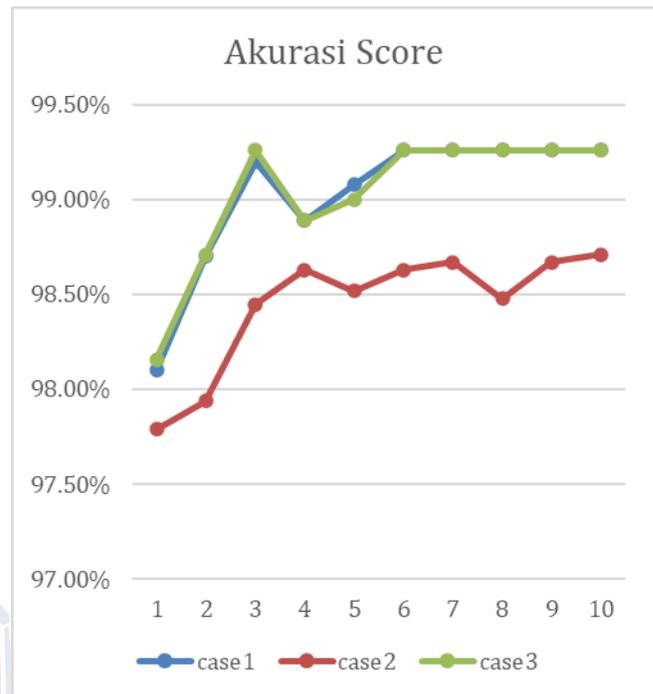
Use case 2:

- Akurasi : 98.44 %
- F1 score : 98.21 %
- Presisi : 87.77 %
- Recall : 57.2 %

Use case 3:

- Akurasi : 99 %
- F1 score : 98.9 %
- Presisi : 64.13 %
- Recall : 60.51 %

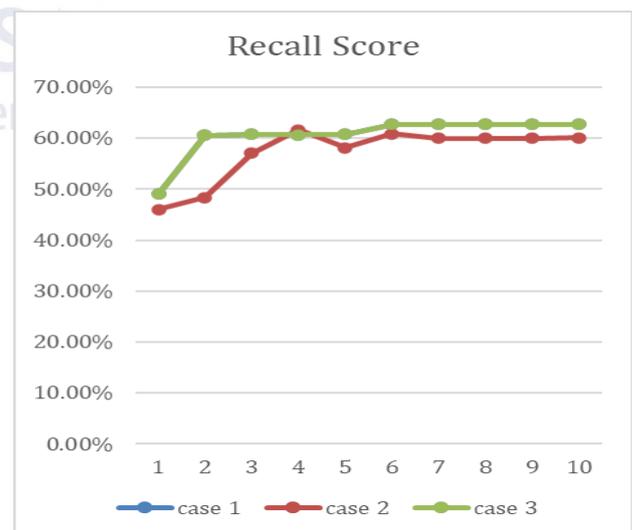
Untuk memudahkan hasil tabel diatas akan dijelaskan secara grafik untuk parameter akurasi:



Gbr. 9 grafik nilai Akurasi

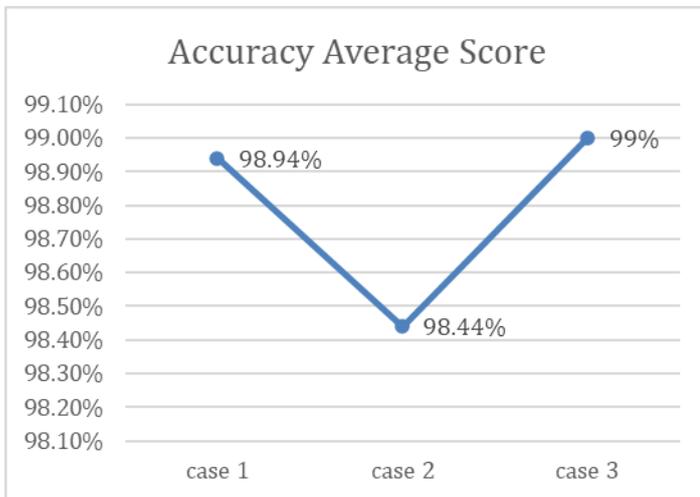
Berdasarkan hasil analisa grafik diatas, didapatkan hasil akurasi yang signifikan dari 3 use case. Use case 1 dengan perbandingan 90:10 data latih dan uji, use case 2 dengan perbandingan 50:50, use case 3 dengan perbandingan 70:30. Berdasarkan hasil diatas use case kedua memiliki hasil terendah dibandingkan use case kesatu dan ketiga.

Parameter kedua yakni Recall yang akan dijelaskan pada grafik dibawah ini:



Gbr.10 Grafik nilai Recall

Berdasarkan hasil grafik diatas diperoleh hasil nilai yang berdekatan antara *case 1* dan *case 3*. Sementara untuk *case* ketiga memiliki nilai dibawah dari *case 1* dan *case 3* hanya saja ketika epoch ke-4 memiliki nilai diatas *case 1* dan *case 3*. Dengan ini *case 3* dan *case 1* memiliki nilai tertinggi pada parameter *recall*.



Gbr. 11 Grafik rata-rata nilai akurasi

Pada hasil rata-rata dan grafik diatas penulis menyimpulkan untuk menggunakan *use case* ketiga dengan proporsi 70:30 dikarenakan memiliki *F1 score* dan akurasi lebih tinggi dengan nilai presisi dan *recall* yang rendah. Hal ini dinilai bahwa proporsi 70:30 tidak menyebabkan model “*overfitting*” yang dalam artian data dinilai terlalu *optimistic* hingga mencapai hampir 100% pada salah satu epoch dengan nilai *loss* yang rendah.

IV KESIMPULAN

Berdasarkan review komentar pengguna Ruang Guru di Google Play Store. *Review* pengguna mayoritas menyatakan positif. Dari 5437 *testing data*, 5254 komentar yang dinyatakan positif, 16 yang dinyatakan netral sentimen, dan 167 komentar dinyatakan negatif sentimen. Disini dapat disimpulkan apabila komentar pengguna Ruang Guru di Google Play Store menyukai aplikasi Ruang Guru berdasarkan jumlah sentimen positif. Berdasarkan dari nilai *F1 Score* yang didapat dari presisi dan *recall* yakni 98.9% dan nilai akurasi bernilai 99%. Dengan ini dapat disimpulkan bahwa metode *pre-trained BERT* sangat efektif untuk di implementasi analisis sentimen.

V SARAN

Berdasarkan penelitian yang telah dilakukan, saran yang dapat diberikan untuk penelitian ini adalah:

1. Konsisten terhadap data yang digunakan
2. Langkah *preprocessing* harus lebih rinci lagi termasuk normalisasi
3. Penerapan algoritma BERT pada platform yang berbeda

REFERENSI

- 1 Ruangguru.(2018). Tentang Ruangguru. Available at: <https://www.ruangguru.com/about-us>, diakses pada 7 Februari 2021.
- 2 Bang, L., & Nitin, J. (2008). Opinion Spam and Analysis. Proceedings of the 2008 International Conference on Web Search and Data Mining, 219–230.
- 3 Devika, M., Sunitha, C., & Amal, G. (2016). Sentiment Analysis:A Comparative Study On Different Approaches. Fourth International Conference on Recent Trends in Computer Science & Engineering, 44 – 49.
- 4 Bharati, M. & Ramageri, Bharati. (2010). Data mining techniques and applications. Indian Journal of Computer Science and Engineering. 1.
- 5 Schäfer, Franziska & Zeiselmaier, Christian & Becker, Jonas & Otten, Heiner. (2018). Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes. 190-195. 10.1109/ITMC.2018.8691266.
- 6 Imtiyas,M.Z., Nasrun,Muhammad, S.Si.,M.T., & Ahmad,Ali U..T.,M.T.(2015). Analisis dan Implementasi *Framework* CRISP-DM Untuk Mengetahui Perilaku Data Transaksi Pelanggan.
- 7 Budi, H., Yova, R., Fatur, R., Julius, D. T., Ruth, M., & Muhammad, Y. F. (2019). Facebook Analysis of Community Sentiment on 2019 Indonesian. The Fifth Information Systems International Conference 2019, 715-722.
- 8 Alifia, R. P., & Irfandy, T. (2020). Sentiment Analysis for Customer Review: Case Study of. Journal of Information System Engineering and Business Intelligence , 1-8.
- 9 Lutfi Anang A., Permanasari ,Adhistya E & Fauziati Silmi, (2018). Department of Electrical Engineering and Information Technology,Universitas Gadjah Mada.
- 10 Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding(2018).
- 11 Saad, A. R., Yan, Q., Syed, B. A., & Waqas, A. (2019). Using BERT for Checking the Polarity of Movie Reviews.
- 12 Thomas W. Edgar, David O. Manz., (2017). In Research Methods for Cyber Security.