

5. PENGUJIAN SISTEM

Pengujian akan dilakukan menggunakan *hyperparameters* yang sama untuk proses *training* model. Karena keterbatasan *resource*, skripsi ini tidak dapat menyertakan *hyperparameter tuning*. *Hyperparameter* yang digunakan adalah *hyperparameter default* dari setiap model. Untuk masing – masing model, berikut merupakan *hyperparameters* yang digunakan :

1. *Pointer-Generator Networks*

- a. *Hidden Dimension* : 256
- b. *Embedding Dimension* : 128 dan 756
- c. *Batch Size* : 32
- d. *Vocabulary Size* : 50000 dan 32000
- e. *Maximum Encoder Steps* : 500
- f. *Beam Size* : 4
- g. *Minimum Decoder Sequence* : 60 dan 100
- h. *Maximum Decoder Sequence* : 100 dan 120
- i. *Learning Rate* : 0.3
- j. *Adagrad Optimizer's Initial Accumulator* : 50000
- k. *Random-Uniform Initialization's Magnitude* : 0.02
- l. *Truncated-Normal Initializer's Standard-Deviation* : 0.0001
- m. *Gradient Clipping* : 2.0
- n. *Coverage Loss Weight* : 1.0

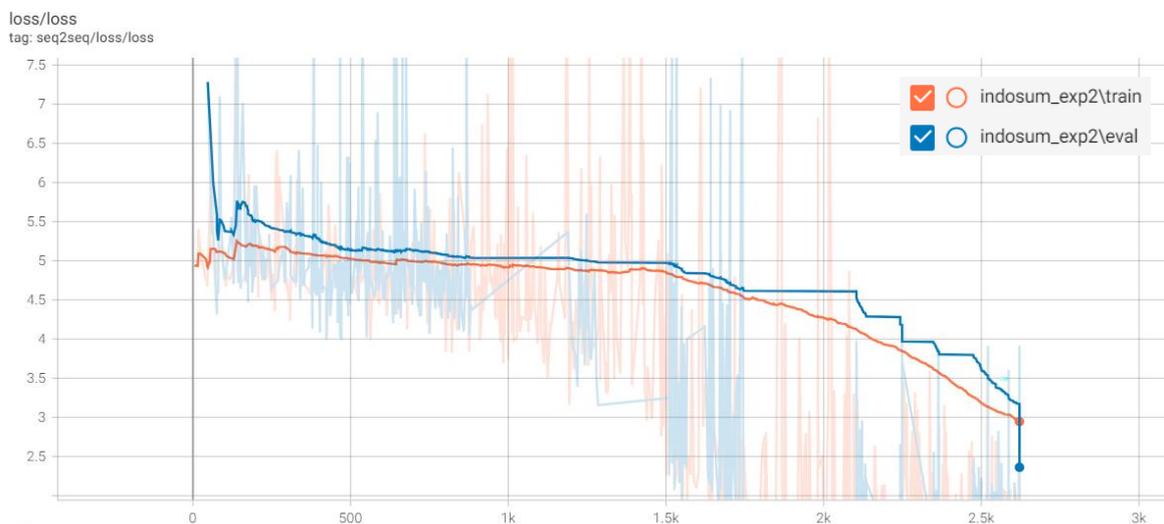
2. *IndoBERT*

- a. *Minimum Length* : 20 dan 40
- b. *Maximum Length* : 80 dan 100
- c. *Number of Beams* : 10
- d. *Repetition Penalty* : 2.5
- e. *Length Penalty* : 1.0
- f. *Early Stopping* : True
- g. *No n-grams Repeat Size* : 2
- h. *Temperature* : 0.8
- i. *Top K* : 50
- j. *Top P* : 0.95

Menggunakan *hyperparameters* diatas, dilakukan proses implementasi sistem menggunakan *device* dengan spesifikasi berikut :

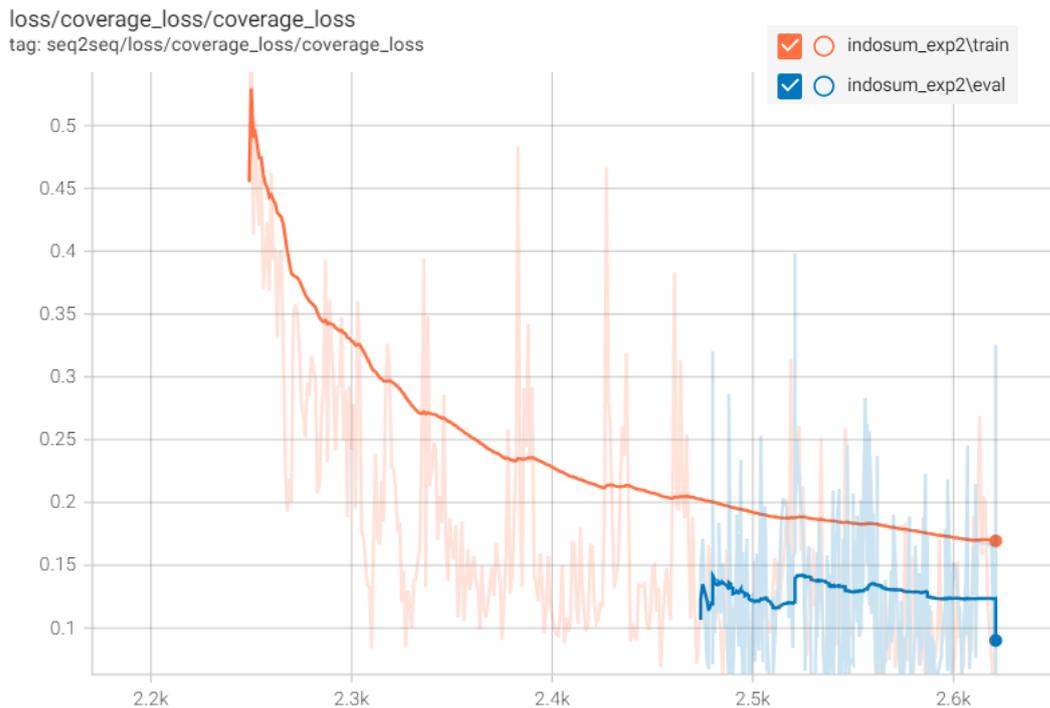
- *Processor* : Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz
- *RAM* : 8 GB
- *OS* : Windows 11, 64-bit
- *Graphic Card* : NVIDIA GeForce GTX 1650 Ti with Max-Q Design

Berikut merupakan visualisasi grafik *loss* yang didapat setelah melalui proses *training* dan *validation*. IndoBERT merupakan *pretrained* model dan tidak dilakukan *fine-tune*, maka grafik dibawah merupakan grafik hasil proses *training* model PGN :



Gambar 5.1. Grafik Loss PGN

Gambar diatas adalah grafik penurunan *loss* pada PGN *baseline*, yaitu dengan Tensorflow Word Embedding. Garis kuning melambangkan penurunan *training loss*, sedangkan garis biru melambangkan penurunan *validation loss*. Proses *validation loss* terkadang diberhentikan (pada area datar) untuk mempercepat proses *training*. Dapat dilihat bahwa grafik *training loss* masih mengalami penurunan, diikuti oleh grafik *validation loss*. Hal ini menandakan bahwa *training* dan *validation loss* masih dapat menurun. Namun, karena keterbatasan *resource*, proses *training* dihentikan pada 2.621 *step* dengan *training loss* 2,947 dan *validation loss* 2,362.



Gambar 5.2. Grafik Coverage Loss PGN

Gambar diatas merupakan grafik penurunan *coverage loss*. Fitur *coverage mechanism* baru diaktifkan pada *step* ke 2.249, sesuai dengan saran dari pencetus PGN yang menganjurkan untuk mengaktifkan *coverage mechanism* pada akhir proses *training*. *Coverage loss* tertinggi selama proses *training* adalah 0,5295 dan mencapai angka 0,1694 pada akhir proses *training*. Pada proses *validation*, *coverage loss* dimulai pada angka 0,1066 dan berakhir pada angka 0,0901 pada akhir proses *validation*.

5.1. Analisis Hasil Ringkasan

Berikut merupakan beberapa *sample* hasil ringkasan dari *Pointer-Generator Networks* dan IndoBERT :

1. Sample 1

- **Artikel**

Lihat artikel pada **Lampiran 1**

- **Ringkasan Referensi**

rio kiswinar memenuhi panggilan penyidik polda metro jaya untuk dimintai keterangan terkait laporannya kepada mario teguh , jum'at (14 / 10 / 2016) . ario datang pada siang hari dengan didampingi oleh pengacaranya ferry h. __amahorseya__ . sebelumnya , ario kiswinar melaporkan mario teguh ke polda metro jaya dengan

tuduhan pencemaran nama baik dan fitnah . mario teguh dalam sebuah stasiun televisi menyatakan bahwa ario kiswinar bukan anak kandungnya .

- **Ringkasan PGN**

ario memenuhi memenuhi panggilan penyidik **resmob polda metro resmob polda metro resmob 2016) 2016) 2016) 2016) 2016) 2016) 2016) 2016) 2016) 2016) 2016)** . ario kiswinar datang sekira pukul . ario kiswinar berkomentar , **hanya tersenyum ya , hanya tersenyum ya , hanya tersenyum** sambil , ” kata ario singkat banyak , hanya tersenyum sambil .

- **Ringkasan PGN + IndoBERT**

ario kiswinar memenuhi panggilan penyidik resmob polda metro jaya. ia mengaku hanya tersenyum ya, hanya tersenyum.

- **Analisis Sample**

Pada ringkasan PGN, terdapat frasa repetitif “resmob polda metro” yang dikoreksi dengan menghilangkan repetisi. Frasa “2016)” juga dihilangkan karena dianggap tidak sesuai dengan kalimat sebelumnya. IndoBERT juga berusaha mengkoreksi frasa “hanya tersenyum ya”, meskipun tidak dapat sepenuhnya dipahami, tapi lebih baik dibanding ringkasan awal.

2. Sample 2

- **Artikel**

Lihat artikel pada **Lampiran 2**.

- **Ringkasan Referensi**

kylian mbappe terus mendapat pengakuan soal kapasitasnya sebagai calon megabintang sepakbola setelah berhasil memenangi golden boy 2017 sebagai pemain eropa terbaik di bawah umur 21 tahun . mbappe mengalahkan finalis lainnya , yakni ousmane dembele dan gabriel jesus . berdasarkan voting dari para jurnalis di seluruh dunia , mbappe mengoleksi **__291__** suara , unggul jauh dari dembele (**!!__149**)dan **__!!** gabriel jesus (**72**) .

- **Ringkasan PGN**

kylian mbappe terus mendapat pengakuan soal kapasitasnya sebagai calon megabintang sepakbola setelah berhasil memenangi golden boy 2017 sebagai pemain eropa terbaik di bawah umur 21 tahun . diumumkan oleh surat kabar italia tuttospo pada senin (**23 / 10**) , penyerang muda paris saint - germain itu mengalahkan finalis

lainnya , yakni ousmane dembele dan gabriel jesus , untuk menyabet penghargaan paling bergengsi bagi pemain muda ini .

- **Ringkasan PGN + IndoBERT**

kylian mbappe terus mendapat pengakuan soal kapasitasnya sebagai pemain eropa terbaik di bawah umur 21 tahun. ia berhasil memenangi golden boy kemarin untuk menyabet penghargaan paling bergengsi bagi pemain muda ini.

- **Analisis Sample**

Pada hasil ringkasan IndoBERT, model berhasil menggabungkan beberapa konteks pada ringkasan PGN dan menjadikan ringkasan final lebih sederhana. Frasa “kapasitasnya sebagai pemain eropa” dihasilkan setelah memindahkan konteks “golden boy” kepada kalimat kedua, yang digabung dengan menggabungkan konteks “penghargaan paling bergengsi ...” sebagai konteks “golden boy”

3. Sample 3

- **Artikel**

Lihat artikel pada **Lampiran 3**.

- **Ringkasan Referensi**

presiden anyar uefa aleksander ceferin telah mengungkapkan logo resmi untuk turnamen euro 2020 mendatang . logo tersebut didesain oleh y&r yang mencantumkan sejumlah siluet suporter berbagai warna yang merefleksikan banyak kultur . berbicara dalam seremoni di london , ceferin mengatakan bahwa uefa menginginkan agar turnamen 2020 tersebut menjadi selebrasi nyata dari sepak bola yang dicintai dan dinikmati .

- **Ringkasan PGN**

apresiden anyar uefa aleksander ceferin telah mengungkapkan logo resmi untuk turnamen euro 2020 mendatang . logo tersebut , yang didesain oleh y&r , mencantumkan sejumlah siluet suporter berbagai warna yang merefleksikan banyak kultur mengingat turnamen tersebut akan digelar di 13 negara eropa . berbicara dalam seremoni di london , ceferin mengatakan : “ uefa ingin agar turnamen 2020 itu menjadi selebrasi nyata dari sepakbola yang kita cintai dan nikmati . “ adakah hal yang lebih baik ketimbang menghadirkan turnamen itu , untuk sekali ini saja , ke seluruh penjuru benua kita yang indah ini ? ” selain london , euro 2020 akan dimainkan di azerbaijan negara eropa .

- **Ringkasan PGN + IndoBERT**

presiden anyar uefa, aleksander ceferin mengungkapkan logo resmi turnamen euro 2020 yang akan digelar di 13 negara eropa. hal ini dimaksudkan agar turnamen tersebut menjadi selebrasi nyata dari sepakbola inggris yang amat digemari.

- **Analisis Sample**

Pada ringkasan PGN, model berhasil meringkas nama – nama negara di Eropa menjadi “negara eropa”. Meskipun kata “azerbajian” tidak dieliminasi, model berhasil mengeliminasi nama negara – negara lain. Pada ringkasan PGN + IndoBERT, frasa “hal ini dimaksudkan agar turnamen tersebut” berhasil memberi jembatan antara kedua kalimat. Kata “turnamen tersebut” juga lebih mendekati kepada “turnamen 2020 tersebut” pada ringkasan referensi. Meskipun seharusnya frasa tersebut merujuk kepada pembahasan logo, namun IndoBERT tidak memasukkan pembahasan desain logo yang menyangkut konteks “selebrasi nyata dari sepak bola yang kita cintai”.

4. Sample 4

- **Artikel**

Lihat artikel pada **Lampiran 4**.

- **Ringkasan Referensi**

eman ahmed abd el aty memiliki berat badan mencapai 500 kilogram sebelum menjalankan operasi di mumbai maret lalu dimana ia mengurangi seperlima dari berat badannya .abd el aty diberi diet cairan khusus selama berada di india yang bertujuan menurunkan berat badan .kini , berat badannya telah turun drastis sebanyak 323 kilogram dalam tiga bulan .sekarang berat badannya tinggal 176,6 kilogram .

- **Ringkasan PGN**

dilansir afp , seorang warga mesir yang dipercaya sebagai wanita terberat di dunia masuk sebuah rumah sakit di abu dhabi Kamis (5 / 5) , dimana ia akan melanjutkan perawatan setelah menjalankan operasi menurunkan badan di india . eman ahmed abd el aty memiliki berat badan mencapai 500 kilogram sebelum menjalankan operasi di mumbai maret lalu dimana ia mengurangi seperlima dari berat badannya . ia menderita berbagai macam masalah kesehatan termasuk penyakit gajah , kondisi yang menyebabkan anggota badan membengkak . abd el aty diberi diet cairan khusus selama berada di india yang bertujuan menurunkan berat badan .

- **Ringkasan PGN + IndoBERT**

seorang wanita di mesir yang dipercaya sebagai wanita terberat di dunia masuk sebuah rumah sakit di abu dhabi kamis (5 / 5). ia akan melanjutkan perawatan setelah menjalani operasi menurunkan **berat** badan di india.

- **Analisis Sample**

Pada ringkasan PGN, terdapat kesalahan kecil pada frasa “menurunkan badan”. Pada ringkasan PGN + IndoBERT, kesalahan ini berhasil diperbaiki dengan menambahkan kata “berat” menjadi frasa yang benar yaitu “menurunkan berat badan”

5. Sample 5

- **Artikel**

Lihat artikel pada **Lampiran 5**.

- **Ringkasan Referensi**

perwakilan jose mourinho membantah tudingan pemerintah spanyol bahwa kliennya telah melakukan penggelapan pajak selama melatih real madrid . pria asal portugal itu dituduh melakukan penggelapan pajak sebesar 3,3 juta euro atau sekitar rp 48,9 miliar antara tahun 2011 dan 2012 . menurut pengacara mou , kliennya telah membayar 26 juta euro pajak selama bekerja tiga tahun di !! __ __ !! real madrid !! __ __ !! antara 2010 dan 2013 .

- **Ringkasan PGN**

perwakilan jose mourinho membantah tudingan pemerintah spanyol bahwa kliennya telah melakukan penggelapan pajak selama melatih real madrid . pria asal portugal itu dituduh melakukan penggelapan pajak sebesar 3,3 juta euro atau sekitar rp 48,9 miliar antara tahun 2011 dan 2012 . **demikian seperti dikutip dari bbc** . " mourinho tidak melaporkan penerimaan pendapatan atas hak citranya , " demikian sebut salah seorang jaksa penuntut . akan tetapi , melalui pengacaranya , mourinho membantah telah melakukan tindakan tersebut **melakukan penggelapan pajak selama melatih real madrid** . **pria asal portugal itu dituduh melakukan penggelapan pajak sebesar 3,3 juta euro atau sekitar rp 48,9 miliar antara tahun 2011 dan 2012 . demikian seperti dikutip dari bbc** .

- **Ringkasan PGN + IndoBERT**

perwakilan jose mourinho membantah tudingan pemerintah spanyol bahwa kliennya telah melakukan penggelapan pajak selama melatih real madrid. pria asal

portugal itu dituduh melakukan hal yang sama, sekitar rp 48,9 miliar antara tahun 2011 dan 2012.

- **Analisis Sample**

Pada ringkasan PGN, seharusnya kalimat “demikian seperti dikutip dari bbc” tidak diperlukan pada ringkasan. Frasa “melakukan penggelapan pajak” sudah ada pada kalimat sebelumnya yang mana tidak perlu diulang. Frasa “selama melatih real madrid” diambil dari kalimat berikutnya yang membahas pajak yang dibayarkan subjek selama melatih real madrid. Kalimat terakhir merupakan pengulangan kalimat kedua dan ketiga.

Pada ringkasan PGN + IndoBERT, terdapat frasa “hal yang sama”. Frasa ini kemungkinan muncul karena pada ringkasan PGN, terdapat pengulangan frasa “melakukan penggelapan pajak”. Informasi pengulangan ini ditangkap oleh IndoBERT dan model mengeluarkan *output* bahwa model telah menangkap informasi 2 kali.

6. Sample 6

- **Article**

Lihat artikel pada **Lampiran 6**.

- **Ringkasan Referensi**

direktur pidana siber bareskrim mabes polri , fadil imran mendukung wakil sekjen pdi perjuangan eriko sotarduga mengenai syarat pembuatan akun media sosial harus berdasarkan identitas pada e - ktp yang sah . sebelumnya , eriko mengusulkan agar pembuatan akun media sosial harus berdasarkan identitas e - ktp yang sah . pengisian nomor e - ktp juga dijadikan syarat agar pemilik akun mudah dilacak .

- **Ringkasan PGN**

direktur pidana siber bareskrim mabes polri , fadil imran mendukung wakil sekjen pdi perjuangan eriko sotarduga mengenai syarat pembuatan akun media sosial harus berdasarkan identitas pada e - ktp yang sah . sebelumnya , eriko mengusulkan agar pembuatan akun media sosial harus berdasarkan identitas e - ktp yang sah . pengisian nomor e - ktp juga dijadikan syarat agar pemilik akun mudah dilacak .

5.2. Pengujian Metrik ROUGE

Metrik ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) sudah menjadi salah satu metrik utama dalam mengevaluasi tugas *Text Summarization*. ROUGE berguna untuk secara

eksplisit menghitung kemiripan kata yang digunakan antara *reference summary* dan *generated summary*.

Pada skripsi ini akan digunakan 3 jenis metrik ROUGE, yaitu ROUGE-1, ROUGE-2, dan ROUGE-L (L merupakan *sequence* terpanjang). ROUGE dihitung menggunakan library *pyrouge* setelah *decoder* menghasilkan hasil ringkasan untuk setiap data *test*. Setiap data *test* akan dikalkulasi ROUGE-1, ROUGE-2, dan ROUGE-L. Setiap metrik ROUGE akan dikalkulasi *precision*, *recall*, dan *f1 score*. Nilai ROUGE secara keseluruhan merupakan rata – rata dari ROUGE setiap data *test*. Berikut merupakan hasil dari pengujian metrik ROUGE :

5.2.1. Perbandingan ROUGE PGN dengan PGN + IndoBERT

Tabel 5.1.

Tabel ROUGE pada Model PGN

	Precision	Recall	F1 Score
ROUGE-1	0.4813	0.7579	0.5871
ROUGE-2	0.4245	0.6724	0.5189
ROUGE-L	0.4744	0.7469	0.5786

Tabel 5.2.

Tabel ROUGE pada PGN + IndoBERT

	Precision	Recall	F1 Score
ROUGE-1	0.7263	0.4226	0.5315
ROUGE-2	0.5956	0.3435	0.4332
ROUGE-L	0.6981	0.4064	0.5110

Model pertama merupakan *baseline* model, yaitu hanya menggunakan *Pointer-Generator Networks* untuk menghasilkan ringkasan final. Model kedua menggunakan *pretrained model* IndoBERT untuk memproses hasil ringkasan PGN menjadi ringkasan final. Maka perlu adanya perbandingan hasil ringkasan yang menggunakan IndoBERT dengan yang tidak menggunakan IndoBERT. Pada tabel Tabel 5.1. dan Tabel 5.2. dapat dilihat bahwa nilai ROUGE pada PGN + IndoBERT cenderung lebih rendah. Hal ini wajar adanya karena IndoBERT akan meringkas dari hasil ringkasan PGN.

Tabel 5.3.

Tabel Perubahan Skor ROUGE

	Precision	Recall	F1 Score
ROUGE-1	+24,50%	-33,53%	-5,56%
ROUGE-2	+17,11%	-32,89%	-8,57%
ROUGE-L	+22,37%	-34,05%	-6,76%

Tabel 5.3. merupakan tabel perubahan rata – rata skor ROUGE terhadap pengaplikasian IndoBERT pada ringkasan PGN. Dapat dilihat secara keseluruhan bahwa *precision* mengalami kenaikan, *recall* mengalami penurunan, dan *f1 score* mengalami penurunan. Pada hasil ringkasan PGN saja, rata – rata panjang ringkasan adalah 69,16 *token*, dan pada hasil ringkasan PGN + IndoBERT, rata – rata panjang ringkasan adalah 23,33 *token*.

5.2.2. Perbandingan Panjang Artikel terhadap ROUGE

Panjang artikel akan dibagi menjadi 3 kategori, yaitu *short*, *medium*, dan *long* dengan perhitungan jumlah *token*. Pembagian panjang artikel menggunakan batas persentil 33 dan 66 dari semua artikel. Batas artikel pendek adalah angka persentil 33 berada pada angka 278 *tokens*, dan batas artikel medium adalah angka persentil 66 berada pada angka 372 *tokens*. Artikel pada data *test* akan dikategorikan lalu dilakukan *metric scoring* menggunakan ROUGE. Berikut merupakan hasil perhitungan metrik ROUGE :

Tabel 5.4.

Tabel ROUGE pada PGN + IndoBERT berdasarkan Panjang Artikel

	Article Length	Precision	Recall	F1 Score
ROUGE-1	Short	0.8571	0.4615	0.6000
	Medium	0.5909	0.2167	0.3171
	Long	0.4615	0.4706	0.4660
ROUGE-2	Short	0.7407	0.3922	0.5128
	Medium	0.4762	0.1695	0.2500
	Long	0.3529	0.3600	0.3564
ROUGE-L	Short	0.8571	0.4615	0.6000
	Medium	0.5909	0.2167	0.3171
	Long	0.4423	0.4510	0.4466

Dari tabel 5.4., dapat dilihat bahwa artikel pendek mendapatkan skor paling tinggi. Perlu diingat bahwa pada hasil eksperimen, *generated summary* cenderung ekstraktif karena *label* dari dataset cenderung mengambil beberapa kalimat pertama (tanpa ada perubahan). Maka, pada nilai *precision* dimana pembandingnya adalah *generated summary*, maka makin pendek sebuah *generated summary*, akan lebih mudah untuk mendapatkan skor presentase *overlapping n-grams* tinggi. Sebaliknya, pembanding dari *recall* merupakan *reference summary*, maka makin panjang sebuah *generated summary*, akan lebih susah untuk mendapatkan skor presentase *overlapping n-grams* tinggi.

5.3. Pengujian Metrik BERTScore

Berbeda dengan ROUGE yang menghitung berdasarkan kata secara eksplisit, metrik BERTScore akan menghitung kemiripan antara *reference summary* dan *generated summary* berdasarkan konteks kalimat. BERTScore berguna untuk mengevaluasi model yang bersifat *hybrid* atau *abstractive* karena kata yang dihasilkan dapat berbeda dengan kata yang ada pada *reference summary* meskipun artinya sama atau mirip.

Pada penelitian sebelumnya tentang PGN Bahasa Indonesia, belum menggunakan metrik BERTScore. Pada skripsi ini akan ditambahkan metrik BERTScore (menggunakan *library bert_score* dengan model BERT 'bert-base-multilingual-cased') karena lebih sesuai untuk mengevaluasi model *hybrid* seperti PGN. BERTScore akan menghasilkan *precision*, *recall*, dan *f1 score*. Nilai BERTScore secara keseluruhan merupakan nilai rata – rata dari BERTScore pada setiap data *test*. Berikut merupakan hasil dari pengujian metrik BERTScore pada model PGN dan PGN + IndoBERT :

5.3.1. Perbandingan BERTScore pada PGN dengan PGN + IndoBERT

Tabel 5.5.

Tabel Metrik BERTScore Overall

	Precision	Recall	F1 Score
PGN	0.7910	0.8443	0.8164
PGN + IndoBERT	0.8480	0.7768	0.8105

Model pertama merupakan *baseline* model, yaitu hanya menggunakan *Pointer-Generator Networks* untuk menghasilkan ringkasan final. Model kedua menggunakan *pretrained model* IndoBERT untuk memproses hasil ringkasan PGN menjadi ringkasan final. Maka perlu

adanya perbandingan hasil ringkasan yang menggunakan IndoBERT dengan yang tidak menggunakan IndoBERT yang dapat dilihat pada Tabel 5.5. diatas.

Tabel 5.6.

Tabel Perubahan Skor BERTScore

	Precision	Recall	F1 Score
BERTScore	+5,70%	-6,75%	-0.59%

Tabel 5.6. merupakan tabel perubahan rata – rata skor BERTScore terhadap pengaplikasian IndoBERT pada ringkasan PGN. Dapat dilihat secara keseluruhan bahwa *precision* mengalami kenaikan, *recall* mengalami penurunan, dan *f1 score* mengalami penurunan yang tidak signifikan. Pada hasil ringkasan PGN saja, rata – rata panjang ringkasan adalah 69,16 *token*, dan pada hasil ringkasan PGN + IndoBERT, rata – rata panjang ringkasan adalah 23,33 *token*.

5.3.2. Perbandingan Panjang Artikel terhadap BERTScore

Panjang artikel akan dibagi menjadi 3 kategori, yaitu *short*, *medium*, dan *long* dengan perhitungan jumlah *token*. Pembagian panjang artikel menggunakan batas persentil 33 dan 66 dari semua artikel. Angka persentil 33 berada pada angka 278, dan angka persentil 66 berada pada angka 372. Artikel pada data *test* akan dikategorikan lalu dilakukan *metric scoring* menggunakan ROUGE. Berikut merupakan hasil perhitungan metrik ROUGE :

Tabel 5.7.

Tabel ROUGE pada PGN + IndoBERT berdasarkan Panjang Artikel

	Article Length	Precision	Recall	F1 Score
BERTScore	Short	0.8581	0.7830	0.8188
	Medium	0.7877	0.7396	0.7629
	Long	0.7831	0.8037	0.7933

Pada tabel 5.7., artikel pendek mendapatkan nilai tertinggi. Sama seperti ROUGE, karena hasil ringkasan cenderung ekstraktif, maka nilai *precision* cenderung lebih tinggi jika *generated summary* lebih pendek, dan nilai *recall* cenderung lebih tinggi rendah jika *generated summary* lebih pendek. Namun pada *f1 score*, tidak terdapat perubahan signifikan. Dari hasil analisa,

BERTScore belum mampu secara jelas memilah mana yang memiliki *hallucination* dengan memberikan skor yang jauh lebih rendah. Skor tetap akan berkurang karena konteks kedua ringkasan berbeda, tapi belum dapat disimpulkan skala ada atau tidaknya *hallucination* pada ringkasan.

5.4. Eksperimen Panjang Ringkasan PGN terhadap Ringkasan Final

Karena ringkasan PGN akan dimasukkan ke model BertSum-IndoBERT, maka panjang ringkasan PGN akan sangat mempengaruhi hasil ringkasan akhir. Jika ringkasan hasil PGN terlalu pendek, ringkasan dapat hanya 1 kalimat yang terdiri dari sedikit kata – kata. Maka perlu adanya perbandingan eksperimen pada panjang ringkasan PGN.

Perubahan yang dilakukan adalah pada *minimum decoder sequence* dan *maximum decoder sequence*. Pada penelitian Abigail See, pencetus metode *Pointer-Generator Networks*, ia menggunakan angka 60 sebagai *minimum decoder sequence* dan 100 sebagai *maximum decoder sequence*. Pada eksperimen ini akan mengganti angka – angka tersebut menjadi 100 sebagai *minimum decoder sequence* dan 120 sebagai *maximum decoder sequence*. Pada eksperimen ini dihindari ringkasan PGN yang terlalu panjang, supaya PGN tidak memaksakan untuk memasukkan bagian yang kurang penting pada ringkasan PGN.

Selain PGN, model IndoBERT juga akan dilakukan *tuning* supaya hasil ringkasan final tidak terlalu pendek ataupun terlalu panjang. Parameter *default* dari IndoBERT adalah menggunakan angka 20 sebagai *minimum decoder sequence*, dan angka 80 sebagai *maximum decoder sequence*. Pada eksperimen ini, parameter tersebut akan diganti menjadi 40 dan 100 sebagai *minimum* dan *maximum decoder sequence*. Berikut merupakan penamaan parameter :

- *Baseline Sequence* :
 - PGN : *Minimum 60, Maximum 100*
 - IndoBERT : *Minimum 20, Maximum 80*
- *Longer Sequence* :
 - PGN : *Minimum 100, Maximum 120*
 - IndoBERT : *Minimum 40, Maximum 100*

Berdasarkan penamaan parameter diatas, akan digabungkan antar model dengan kombinasi sebagai berikut :

1. *Baseline + Baseline Sequence*
2. *Longer + Baseline Sequence*
3. *Longer + Longer Sequence*

Setiap kombinasi akan dilakukan evaluasi metrik ROUGE dan BERTScore.

5.4.1. Perubahan Panjang Ringkasan pada PGN

Tabel 5.8. dibawah merupakan hasil dari evaluasi metrik ROUGE dan BERTScore dengan perubahan *sequence length* pada PGN.

Tabel 5.8.

Tabel Skor Metrik pada PGN dengan Panjang Ringkasan

	ROUGE-1			ROUGE-2			ROUGE-L			BERT Score		
	P	R	F	P	R	F	P	R	F	P	R	F
Baseline	67,00	68,39	67,35	59,84	61,05	60,13	66,09	67,45	66,44	85,29	85,30	85,27
Longer	48,13	75,79	58,71	42,45	67,24	51,89	47,44	74,69	57,86	79,10	84,43	81,64

5.4.2. Perubahan Panjang Ringkasan pada PGN + IndoBERT

Tabel 5.9. dibawah merupakan hasil dari evaluasi metrik ROUGE dan BERTScore dengan perubahan *sequence length* pada PGN + IndoBERT.

Tabel 5.9.

Tabel Skor Metrik pada PGN + IDB dengan Panjang Ringkasan

	ROUGE-1			ROUGE-2			ROUGE-L			BERT Score		
	P	R	F	P	R	F	P	R	F	P	R	F
Base + Base	79,59	31,84	44,78	66,95	26,25	37,07	77,46	31,00	43,59	85,51	74,02	79,29
Long + Base	72,63	42,46	53,15	59,56	34,35	43,32	69,81	40,64	51,10	84,80	77,68	81,05
Long + Long	72,78	42,37	53,28	59,70	34,45	43,44	67,30	39,23	49,31	84,85	77,72	81,10

5.4.3. Analisis Hasil Ringkasan terhadap Perubahan Sequence Length

Tabel 5.10.

Tabel Panjang Ringkasan

PGN			IndoBERT		
Seq. Length	# sentences	# tokens	Seq. Length	# sentences	# tokens
Baseline	3.1696	69.1616	Baseline	1.5808	23.3301
Longer	4.8184	106.4949	Baseline	2.3227	34.4181
			Longer	2.3283	34.4343

Tabel 5.10. merupakan perbandingan panjang ringkasan terhadap perubahan *minimum* dan *maximum decoder sequence*. Dapat dilihat bahwa perubahan parameter *maximum* dan *minimum sequence* untuk model IndoBERT tidak memberi dampak yang signifikan terhadap panjang hasil ringkasan. Pada tabel diatas, model kombinasi *longer + longer* memperoleh rata – rata *f1 score* terbaik dibanding kedua kombinasi lain meskipun perbedaan skor sangat kecil.

5.6. Analisis Out of Context Error (Hallucination)

Hallucination menjadi salah satu masalah utama dalam tugas *text summarization* terutama pada model *summarization* yang bersifat *abstractive* ataupun *hybrid*. *Pointer-Generator Networks* merupakan model dengan jenis *hybrid*, model dapat memilih kapan sebaiknya mengambil *token* dari artikel asli (*pointer mechanism*) dan kapan sebaiknya memilih *token* dari *vocabulary* (*generative mechanism*). Berikut merupakan beberapa *sample* yang terdeteksi adanya *hallucination* :

1. Sample 1

- **Artikel**

Lihat artikel pada **Lampiran 7**.

- **Ringkasan Referensi**

bourne-mouth mampu membuat malu chelsea di depan pendukungnya dengan membantai 3 - 0 pada laga lanjutan liga inggris di stampford bridge kamis (1 / 2 / 2018) dinihari wib . akibat kekalahan ini , chelsea turun ke posisi keempat dengan mengemas nilai 50 sama dengan liverpool di atasnya yang unggul selisih gol . dan kekalahan ini membuat chelsea kian menjauh dari manchester city yang kokoh di puncak klasemen .

- **Ringkasan (PGN)**

inggris– luar biasa bourne-mouth mampu membuat malu chelsea di depan pendukungnya dengan membantai juara bertahan liga inggris itu 3 - 0 tanpa balas . pada laga lanjutan liga inggris yang helat stampford bridge kamis (1 / 2 / 2018) dinihari wib tadi bourne-mouth bourne-mouth unggul lewat gol wilson menit ke - 51 , gol stanislas menit ke - 64 dan gol ake menit ke - 67 .

Tabel 5.11.

Tabel Metrik Sample 1 pada PGN dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	35,29	48,39	40,82
ROUGE-2	25,00	34,43	28,97

ROUGE-L	32,94	45,16	38,10
BERTScore	71,66	77,14	74,30

- **Ringkasan (PGN + IndoBERT)**

inggris membantai juara bertahan liga **premier**, chelsea, 3 - 0 (**1 - 2**) dalam laga lanjutan premiership yang digelar dinihari tadi.

Tabel 5.12.

Tabel Metrik Sample 1 pada PGN + IDB dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	29,63	12,90	17,98
ROUGE-2	7,69	3,28	4,60
ROUGE-L	22,22	9,68	13,48
BERTScore	68,26	65,78	67,00

- **Analisis Sample**

Pada *sample* ini dapat ditemukan *hallucination*. Kata "inggris" pada ringkasan PGN merupakan sumber berita atau informasi topik artikel menyangkut negara tertentu yang gagal dihilangkan oleh PGN. Namun IndoBERT mengartikannya sebagai **subjek**. Kata "premier" merupakan hasil *generative* dari IndoBERT yang *valid* (**premiere league** merupakan sebutan liga inggris). Kata "(1 - 2)" merupakan *hallucination* yang mungkin diambil dari **tanggal pertandingan** (1 / 2 / 2018).

2. Sample 2

- **Artikel**

Lihat artikel pada **Lampiran 8**.

- **Ringkasan Referensi**

presiden joko widodo dijadwalkan akan melantik sejumlah orang dalam waktu dekat ini . setidaknya , jokowi rencananya akan melantik gubernur dki jakarta definitif djarot saiful hidayat , badan siber dan sandi negara (bssn) dan unit kerja presiden pembinaan ideologi pancasila (ukp - pip) . jokowi berharap dapat segera melakukan pelantikan . jokowi menyampaikan ia belum mau menyebut siapa saja personel ukp - pip yang akan segera dilantik .

- **Ringkasan (PGN)**

presiden joko widodo dijadwalkan akan melantik sejumlah orang dalam waktu dekat ini . setidaknya , jokowi rencananya akan melantik gubernur dki jakarta definitif

djarot saiful hidayat , badan siber dan sandi negara (bssn) dan unit kerja presiden pembinaan ideologi pancasila (ukp - pip) . menurut jokowi , ia berharap dapat segera melakukan pelantikan . hal itu dikatakan jokowi usai buka bersama dengan pimpinan lembaga tinggi negara di rumah dinas ketua dpr setya novanto di jalan widya chandra iii , jakarta , senin (5 / 6) .

Tabel 5.13.

Tabel Metrik Sample 2 pada PGN dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	62,67	78,33	69,63
ROUGE-2	56,76	71,19	63,16
ROUGE-L	61,33	76,67	68,15
BERTScore	82,79	86,24	84,48

- **Ringkasan (PGN + IndoBERT)**

presiden joko widodo dijadwalkan akan **melantik sejumlah gubernur** dki jakarta definitif. menurut **presiden**, ia berharap dapat segera melakukan pelantikan.

Tabel 5.14.

Tabel Metrik Sample 2 pada PGN + IDB dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	94,74	30,00	45,57
ROUGE-2	72,22	22,03	33,77
ROUGE-L	89,47	28,33	43,04
BERTScore	87,44	72,61	79,34

- **Analisis Sample**

Pada *sample* ini, terdapat *hallucination* pada frase “melantik sejumlah gubernur”. Kemungkinan kesalahan ini karena IndoBERT berusaha meringkas dengan menggabungkan frasa “melantik sejumlah orang” dan “melantik gubernur” pada kalimat setelahnya. IndoBERT gagal melihat kalimat secara utuh. Berikutnya, kata “joko widodo” diganti oleh IndoBERT menjadi kata “presiden” pada frase “menurut presiden”.

3. Sample 3

- **Artikel**

Lihat artikel pada **Lampiran 9**.

- **Ringkasan Referensi**

setelah terdegradasi ke liga 3 , persikabo bogor tengah menghadapi ancaman pembekuan dari kompetisi . ancaman ini datang setelah pengadilan fifa di zurich , swiss , memutuskan persikabo bersalah atas sengketa mantan pemainnya , jacques joel tsimi . permasalahan berupa klaim joel tsimi yang merasa gajinya belum dibayarkan persikabo . pemain asal kamerun itu menjadi bagian dari laskar pajajaran pada musim 2013 - 2014 saat berlaga di divisi utama .

- **Ringkasan (PGN)**

setelah terdegradasi ke liga 3 , persikabo bogor tengah menghadapi ancaman pembekuan dari kompetisi . ancaman ini datang setelah pengadilan fifa di zurich , swiss , memutuskan persikabo bersalah atas sengketa mantan pemainnya , jacques joel tsimi . permasalahan itu berupa klaim joel tsimi yang merasa gajinya belum dibayarkan persikabo . pemain asal kamerun itu menjadi bagian dari laskar pajajaran pada musim 2013 - 2014 saat berlaga di divisi utama .

Tabel 5.15.

Tabel Metrik Sample 3 pada PGN dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	1,96	1,59	1,75
ROUGE-2	0	0	0
ROUGE-L	1,96	1,59	1,75
BERTScore	60,17	61,33	60,75

- **Ringkasan (PGN + IndoBERT) – Baseline + Baseline**

persikabo bogor **terancam dibekukan** dari kompetisi. **tindakan ini terkait dengan** sengketa mantan pemainnya, jacques joel tsimi.

Tabel 5.16.

Tabel Metrik Sample 3 pada PGN + IDB (Base + Base) dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	17,39	7,55	10,53
ROUGE-2	4,54	1,92	2,70
ROUGE-L	8,70	3,77	5,26
BERTScore	69,79	67,76	68,76

- **Ringkasan (PGN + IndoBERT) – Longer + Longer**

persikabo bogor **terancam dibekukan** dari kompetisi. **pasalnya**, pengadilan fifa memutuskan pemain asal kamerun, **jacques joel tsimi bersalah** atas sengketa mantan pemainnya, **michel malmi**. permasalahan itu berupa gajinya belum dibayarkan.

Tabel 5.17. Tabel Metrik Sample 3 pada PGN + IDB (Long + Long) dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	82,76	39,34	53,33
ROUGE-2	46,43	21,67	29,55
ROUGE-L	58,62	27,87	37,78
BERTScore	81,67	75,48	78,45

- **Analisis Sample**

Pada ringkasan PGN + IndoBERT dengan kombinasi *baseline + baseline sequence*, meskipun ringkasan terlalu pendek dan tidak memuat alasan dari permasalahan pada topik tersebut, ringkasan tidak mengandung *hallucination*. Tapi saat menggunakan *longer + longer sequence*, terdapat *hallucination*. Pada topik artikel maupun ringkasan PGN, seharusnya “persikabo bogor” yang diputuskan bersalah terhadap “jacques joel tsimi”, tapi pada hasil ringkasan, **“jacques joel tsimi” yang bersalah** atas **“michel malmi” yang bahkan tidak ada** pada artikel asli maupun ringkasan PGN. Namun, IndoBERT berhasil mengeluarkan **frasa baru yang valid** yaitu “terancam dibekukan” untuk menggantikan “menghadapi ancaman”, dan frasa “tindakan ini terkait dengan...” dan “pasalnya” untuk menggantikan frasa “ancaman ini datang setelah”.

4. Sample 4

- **Artikel**

Lihat artikel pada **Lampiran 10**.

- **Ringkasan Referensi**

apple bakal mulai memasarkan imac pro pada tanggal 14 desember besok . yang bakal dipasarkan lebih dulu adalah varian dengan prosesor intel xeon 8 - core dan 10 - core . keduanya bisa dikonfigurasi dengan ram ecc ddr4 hingga 128 gb , ssd sampai 4 tb dan kartu grafis amd radeon vega 64 16 gb . harganya dimulai di angka \$ 4.999 untuk konfigurasi 8 - core yang terendah .

- **Ringkasan (PGN)**

pc **pc** untuk gamer dan konsumen secara umum , mac untuk kalangan profesional ; kurang lebih seperti ini anggapan sederhana mayoritas orang yang saya

kenal . kalau melihat ekosistem aplikasi yang tersedia , pendapat ini memang ada benarnya ; aplikasi macam final cut pro dan logic pro buatan apple selama ini kerap menjadi pilihan banyak studio film dan musik kenamaan . *kalangan profesional tentu saja mendambakan performa kelas workstation .sayangnya , terakhir kali apple merilis komputer yang memenuhi standar ini adalah di tahun 2013 .kendati demikian , akhir 2017 ini sepertinya bisa menjadi babak baru bagi platform macos dengan kehadiran imac pro .*

*) Catatan : teks yang diberi *italic & underline* adalah ringkasan tambahan dari PGN *longer sequence*

Tabel 5.18.

Tabel Metrik Sample 4 pada PGN dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	14,54 / 13,83	12,90 / 20,97	13,68 / 16,67
ROUGE-2	0 / 1,07	0 / 1,64	0 / 1,30
ROUGE-L	5,46 / 5,32	4,84 / 8,06	5,13 / 6,41
BERTScore	65,74 / 65,88	62,47 / 65,09	64,06 / 65,48

- Ringkasan (PGN + IndoBERT) – Baseline + Baseline

banyak orang yang mengenal apple ipad sebagai pilihan studio film dan musik kenamaan. **namun tahukah anda, jika melihat ekosistem aplikasi tersebut di dunia?**

Tabel 5.19.

Tabel Metrik Sample 4 pada PGN + IDB (Base + Base) dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	17,39	6,45	9,41
ROUGE-2	0	0	0
ROUGE-L	13,04	4,84	7,06
BERTScore	66,51	59,97	63,12

- Ringkasan (PGN + IndoBERT) – Longer + Longer

apple merilis software final cut pro dan logic pro di tahun 2013. **sayangnya,** **hingga kini aplikasi ini kerap menjadi pilihan banyak studio film dan musik kenamaan.** **sayang,** **akhir tahun** ini sepertinya bisa menjadi babak baru bagi platform tersebut.

Tabel 5.20.

Tabel Metrik Sample 4 pada PGN + IDB (Long + Long) dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	18,42	11,29	14,00
ROUGE-2	0	0	0
ROUGE-L	13,16	8,06	10,00
BERTScore	70,64	63,97	67,14

- **Analisis Sample**

Pada ringkasan PGN, kata “pc” kedua kemungkinan berasal dari PGN yang **berusaha mengganti** kata “windows”. Banyak kesalahan yang ada pada ringkasan PGN + IndoBERT *baseline + baseline*. Pertama, frasa “banyak orang yang mengenal apple ipad” terdapat gabungan dari beberapa *error*. Kata **“banyak orang yang mengenal”** berasal dari “mayoritas orang yang saya kenal”, kata **“apple ipad”** berasal dari “final cut pro dan logic pro”. Kedua, frasa “namun tahukah anda, jika melihat ekosistem aplikasi tersebut di dunia?” **tidak dapat dijelaskan asalnya**, kecuali frasa “ekosistem aplikasi”.

Berbeda dengan *baseline + baseline*, PGN + IndoBERT dengan *longer + longer sequence* dapat dikatakan berhasil merangkum ringkasan PGN. Frasa “apple merilis software final cut pro dan logic pro di tahun 2013” berhasil **merangkum 2 kalimat menjadi 1 kalimat** dengan interpretabilitas yang baik. Kalimat kedua dimulai dengan kata “sayangnya” yang mana membuat **arti kalimat menjadi sedikit berbeda**. Kalimat “sayangnya , terakhir kali apple merilis komputer yang memenuhi standar ini adalah di tahun 2013” yang artinya penulis menyayangkan Apple tidak mengeluarkan aplikasi baru, menjadi “sayangnya, hingga kini aplikasi ini kerap menjadi pilihan banyak studio film dan musik kenamaan” yang artinya penulis menyayangkan tidak ada aplikasi baru yang dapat menggantikan software Apple tersebut. Berikutnya, kata “sayang” pada kalimat ketiga **tidak seharusnya digunakan**, dan frasa “akhir tahun ini” **berhasil menggantikan frasa** “akhir 2017 ini”.

5. Sample 5

- **Artikel**

Lihat artikel pada **Lampiran 11**.

- **Ringkasan Referensi**

menurut pakar kemanan cyber pratama persadha , penting bagi pemerintah untuk mendorong developer untuk membuat media sosial atau aplikasi chat lokal . hal terseut disampaikannya setelah maraknya kasus predator anak dalam media sosial .

pemerintah lebih bisa bertindak tegas bila terjadi pelanggaran maupun kejahatan hukum di tanah air serta memudahkan dalam memblokir konten bila terjadi pelanggaran .

- **Ringkasan (PGN)**

tanah air digegerkan oleh penangkapan sejumlah pemakai facebook yang menggunakannya sebagai ajang berbagi dan berburu anak bawah umur untuk memuaskan hasrat seks mereka . polisi setidaknya sudah menangkap 5 tersangka yang diduga menjadi admin grup facebook pedofilia tersebut . para tersangka diancam pasal prostitusi anak , dijerat pasal 27 ayat 1 juncto pasal 45 ayat 1 uu ri nomor 19 tahun 2016 tentang perubahan atas uu nomor 11 tahun 2008 tentang ite .

Tabel 5.21.

Tabel Metrik Sample 5 pada PGN dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	1,96	1,59	1,75
ROUGE-2	0	0	0
ROUGE-L	1,96	1,59	1,75
BERTScore	60,17	61,33	60,75

- **Ringkasan (PGN + IndoBERT)**

tanah air digegerkan oleh penangkapan sejumlah pemakai facebook yang menggunakannya sebagai ajang berbagi dan berburu anak bawah umur untuk memuaskan hasrat seks mereka. Pada Tabel 5.22 dijelaskan...

Tabel 5.22.

Tabel Metrik Sample 5 pada PGN + IDB dengan Hallucination

	Precision	Recall	F1 Score
ROUGE-1	17,39	7,55	10,53
ROUGE-2	4,54	1,92	2,70
ROUGE-L	8,70	3,77	5,26
BERTScore	69,79	67,76	68,76

Terkadang saat P_{gen} lebih condong untuk *generate token*, dan jika pada saat itu terdapat kesalahan pada *vocabulary distribution*, maka *token* yang akan digenerate akan tidak sesuai dan berdampak pada *token – token* berikutnya, hal ini yang disebut *hallucination*.

Karena label dari dataset yang digunakan bersifat ekstraktif, model PGN menjadi sangat ekstraktif dan hanya mengambil 2-3 kalimat pertama dari artikel asli. Hal ini mengakibatkan tidak ada *hallucination* yang terjadi pada PGN. Namun, saat IndoBERT diaplikasikan pada hasil ringkasan PGN, beberapa *sample* terdeteksi adanya *hallucination*.

5.7. Eksperimen IndoBERT sebagai Word Embedding

Pada metode PGN asli menggunakan *trainable word embedding* dari Tensorflow. Pada skripsi ini akan dilakukan eksperimen dengan mengganti *trainable word embedding* dengan *pretrained word embedding* menggunakan IndoBERT. BERT sangat baik dalam memahami sebuah kata secara *semantic*, maka hal ini dilakukan untuk melihat apakah dengan menggunakan BERT *Embedding* hasil ringkasan dapat lebih baik. Versi IndoBERT yang digunakan diambil dari <https://huggingface.co/cahya/bert-base-indonesian-522M>, merupakan IndoBERT yang digunakan untuk model sumarisasi IndoBERT (tanpa dilanjutkan proses *train* untuk tugas *text summarization*).

Eksperimen ini dilakukan dengan mengganti *initial weight* dari *word embedding layer* pada model PGN yang awalnya menggunakan *truncated normal distribution*. Karena keterbatasan *resources*, jumlah *vocabulary* akan dikurangi dari 50.000 menjadi 32.000 *token*. Dengan mengganti *word embedding*, ukuran *embedding* yang awalnya 128 menjadi 756, disesuaikan dengan IndoBERT *embedding*. Karena ukuran *embedding* yang meningkat, proses *training* jauh lebih membutuhkan waktu. Oleh sebab itu proses *training* belum dapat menyentuh titik optimal.

Tabel 5.23.

Perbandingan Penggunaan IndoBERT sebagai Word Embedding

		ROUGE-1			ROUGE-2			ROUGE-L			BERT Score		
		P	R	F	P	R	F	P	R	F	P	R	F
TF Emb	PGN	67,00	68,39	67,35	59,84	61,05	60,13	66,09	67,45	66,44	85,29	85,30	85,27
	+ID B	79,59	31,84	44,78	66,95	26,25	37,07	77,46	31,00	43,59	85,51	74,02	79,29
IDB Emb	PGN	59,56	64,87	61,61	52,25	56,87	54,02	54,69	61,51	58,43	82,26	82,92	82,55
	+ID B	78,18	30,99	43,67	65,15	25,28	35,78	73,48	29,14	41,06	84,99	73,60	78,82

Pada tabel 5.23., dapat dilihat bahwa menerapkan IndoBERT pada *word embedding* belum dapat menghasilkan nilai metrik yang lebih baik. Hal ini mungkin terjadi karena proses *training* yang belum selesai.

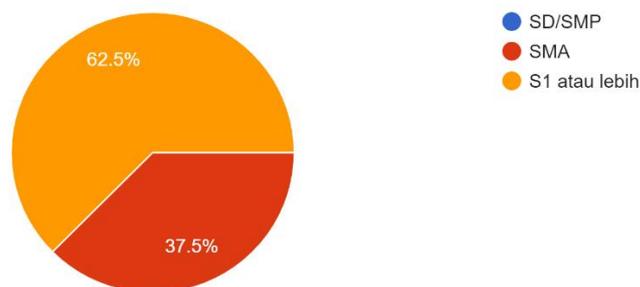
5.8. Survey Penilaian Hasil Ringkasan

Summarization Model merupakan tugas yang sangat subjektif. Beberapa orang mungkin menganggap kalimat tertentu adalah penting, yang mana belum tentu orang lain menganggapnya penting. Dengan begitu, perlu adanya responden yang membantu analisis hasil ringkasan model supaya hasil ringkasan terverifikasi oleh beberapa orang.

Karena target audiens dari artikel adalah masyarakat umum, maka *survey* ini dibagikan kepada masyarakat umum dengan menanyakan pendidikan terakhir dan seberapa sering responden membaca berita dalam 1 minggu. Lalu responden akan diberikan 3 *sample* artikel dan ringkasannya. Responden akan memberikan *rating* dengan beberapa pertanyaan.

Pendidikan Terakhir

16 responses

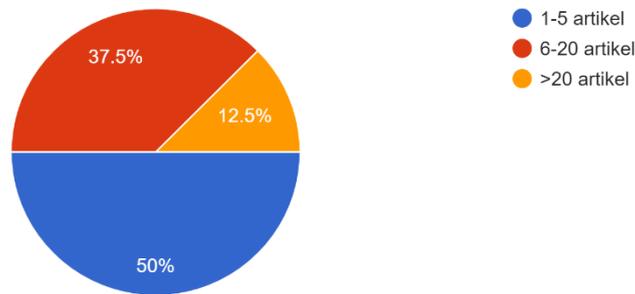


Gambar 5.3. Pie Chart Pendidikan Terakhir

Total responden adalah 16 orang dengan rata – rata responden berpendidikan S1 atau lebih.

Seberapa sering anda membaca berita dalam 1 minggu?

16 responses



Gambar 5.4. Pie Chart Intensitas Membaca Berita

Dalam jangka waktu 1 minggu, rata – rata responden sebanyak 50% membaca 1 hingga 5 artikel, diikuti oleh 37,5% responden membaca 6-20 artikel, dan 12,5% responden membaca lebih dari 20 artikel dalam 1 minggu.

Berikut merupakan beberapa *sample* yang diberikan :

1. *Sample 1*

• **Artikel Berita**

Kylian Mbappe terus mendapat pengakuan soal kapasitasnya sebagai calon megabintang sepakbola setelah berhasil memenangi Golden Boy 2017 sebagai pemain Eropa terbaik di bawah umur 21 tahun. Diumumkan oleh surat kabar Italia Tuttosport pada Senin (23/10), penyerang muda Paris Saint-Germain itu mengalahkan finalis lainnya, yakni Ousmane Dembele dan Gabriel Jesus, untuk menyabet penghargaan paling bergengsi bagi pemain muda ini. Berdasarkan voting dari para jurnalis di seluruh dunia yang dihimpun Tuttosport, Mbappe berada di urutan teratas dengan 291 suara, unggul jauh dari peringkat kedua Dembele (149). Jesus sebetulnya diumumkan Tuttosport masuk tiga besar finalis, namun dalam perhitungan akhir striker Manchester City itu kalah suara (72) dari penyerang Manchester United Marcus Rashford (76). Gianluigi Donnarumma (AC Milan), Christian Pulisic (Borussia Dortmund), Kasper Dolberg (Ajax Amsterdam), Emre Mor (Celta Vigo), Federico Chiesa (Fiorentina), dan Rodrigo Bentancur (Juventus) berturut-turut melengkapi sepuluh besar daftar ini. Dengan keberhasilan memenangi Golden Boy 2017, maka Mbappe mengikuti jejak para pemenang edisi sebelumnya seperti Lionel Messi (2005), Paul

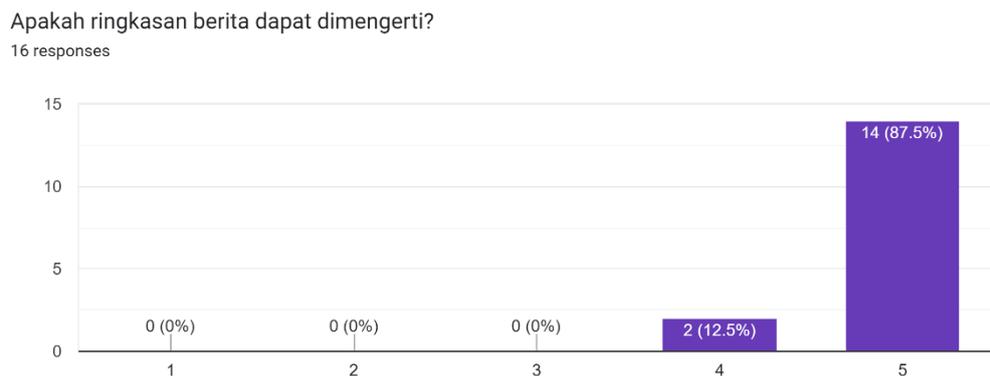
pogba (2013) , anthony martial (2015) , dan renato sanches (2016) . mbappe juga menjadikan prancis sebagai negara penyuplai terbanyak pemenang golden boy , yakni tiga kali (mbappe , pogba , dan martial) . mbappe sebetulnya masih berstatus sebagai pemain pinjaman dari as monaco , namun psg akan mempermanenkannya senilai € 180 juta pada tahun depan untuk __menahbiskannya__ sebagai pemain termahal di dunia kedua di belakang rekan setimnya , neymar . sejauh ini , penyerang 18 tahun itu sudah mengemas empat gol dan empat assist bagi psg di semua kompetisi musim ini .

- **Ringkasan Berita**

kylian mbappe terus mendapat pengakuan soal kapasitasnya sebagai pemain eropa terbaik di bawah umur 21 tahun. ia berhasil memenangi golden boy kemarin untuk menyabet penghargaan paling bergengsi bagi pemain muda ini.

- **Pertanyaan**

a. Apakah ringkasan berita dapat dimengerti?



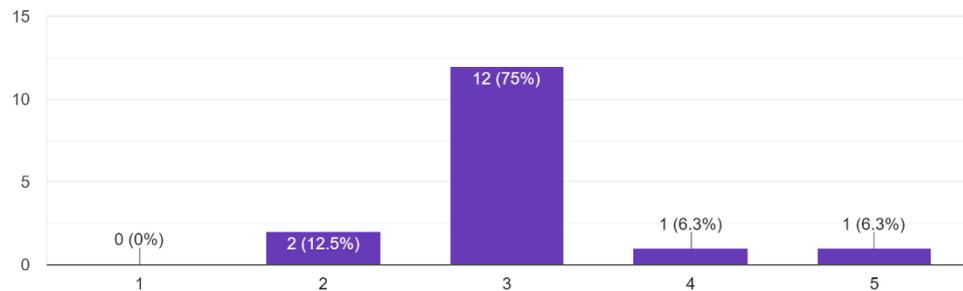
Gambar 5.5. Pertanyaan 1 Sample 1

Pilihan *Rating* adalah 1 (tidak dapat dimengerti) sampai 5 (dapat dimengerti). Dari 16 responden, 87,5% memberi skor 5 dan 12,5% memberi skor 4.

b. Apakah informasi yang dimasukkan ke dalam ringkasan sudah cukup? Note : Ringkasan yang baik mempertahankan inti sari dan menyingkirkan yang tidak penting.

Apakah informasi yang dimasukkan ke dalam ringkasan sudah cukup?

16 responses



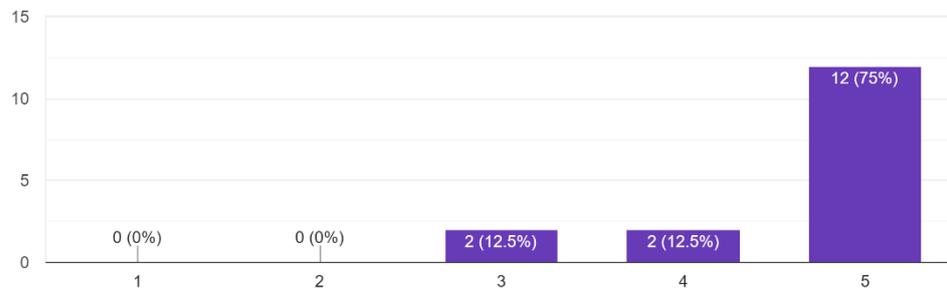
Gambar 5.6. Pertanyaan 2 Sample 1

Pilihan *Rating* adalah 1 (terlalu sedikit informasi) sampai 5 (terlalu banyak informasi) dengan skor paling baik ada 3 (cukup). Dari 16 responden, 75% memberi skor 3 dan 12,5% memberi skor 2.

- c. Apakah ringkasan yang dihasilkan relevan (tidak keluar dari konteks berita) dengan artikel berita?

Apakah ringkasan yang dihasilkan relevan (tidak keluar dari konteks berita) dengan artikel berita?

16 responses



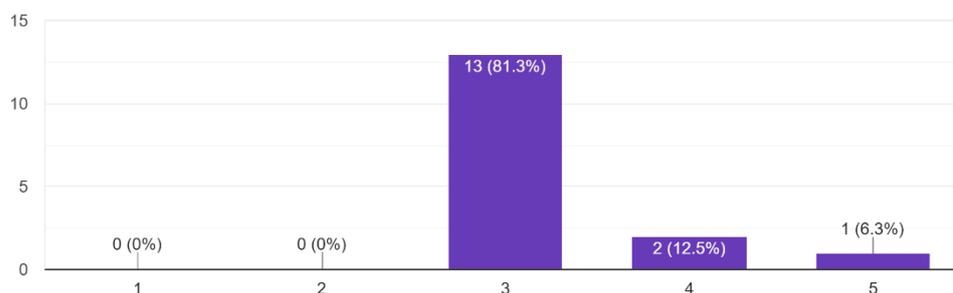
Gambar 5.7. Pertanyaan 3 Sample 1

Pilihan *Rating* adalah 1 (tidak relevan) sampai 5 (relevan). Dari 16 responden, 75% memberi skor 5 dan skor 3 dan 4 diberikan oleh 12,5% responden.

- d. Apakah panjang ringkasan yang dihasilkan sudah cukup?

Apakah panjang ringkasan yang dihasilkan sudah cukup?

16 responses



Gambar 5.8. Pertanyaan 4 Sample 1

Pilihan *Rating* adalah 1 (terlalu pendek) sampai 5 (terlalu panjang) dengan skor terbaik adalah 3 (cukup). Dari 16 responden, 81,3% memberi skor 3 dan 12,5% memberi skor 4.

2. *Sample 2*

• **Artikel Berita**

presiden anyar uefa aleksander ceferin telah mengungkapkan logo resmi untuk turnamen euro 2020 mendatang . logo tersebut , yang didesain oleh __y&r__ , mencantumkan sejumlah siluet suporter berbagai warna yang merefleksikan banyak kultur mengingat turnamen tersebut akan digelar di 13 negara eropa . berbicara dalam seremoni di london , ceferin mengatakan : “ uefa ingin agar turnamen 2020 itu menjadi selebrasi nyata dari sepakbola yang kita cintai dan nikmati . “ adakah hal yang lebih baik ketimbang menghadirkan turnamen itu , untuk sekali ini saja , ke seluruh penjuru benua kita yang indah ini ? ” selain london , euro 2020 akan dimainkan di azerbaijan , belgia , denmark , jerman , hongaria , republik irlandia , italia , belanda , rumania , rusia , skotlandia dan spanyol .

• **Ringkasan Berita**

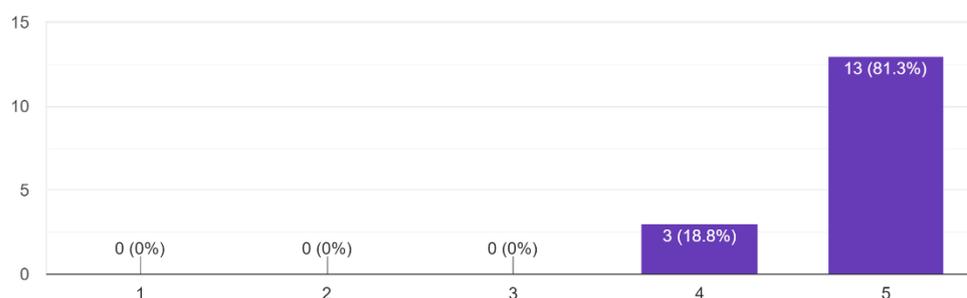
presiden anyar uefa, aleksander ceferin mengungkapkan logo resmi turnamen euro 2020 yang akan digelar di 13 negara eropa. hal ini dimaksudkan agar turnamen tersebut menjadi selebrasi nyata dari sepakbola inggris yang amat digemari.

• **Pertanyaan**

a. Apakah ringkasan berita dapat dimengerti?

Apakah ringkasan berita dapat dimengerti?

16 responses



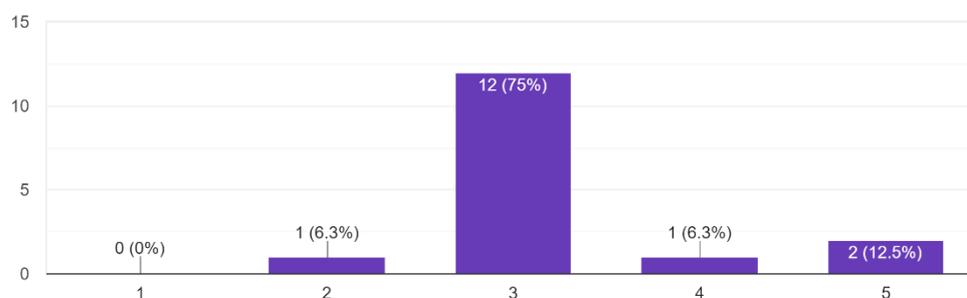
Gambar 5.9. Pertanyaan 1 Sample 2

Pilihan *Rating* adalah 1 (tidak dapat dimengerti) sampai 5 (dapat dimengerti). Dari 16 responden, 81,3% memberi skor 5 dan 18,8% memberi skor 4.

- b. Apakah informasi yang dimasukkan ke dalam ringkasan sudah cukup? Note : Ringkasan yang baik mempertahankan inti sari dan menyingkirkan yang tidak penting.

Apakah informasi yang dimasukkan ke dalam ringkasan sudah cukup?

16 responses

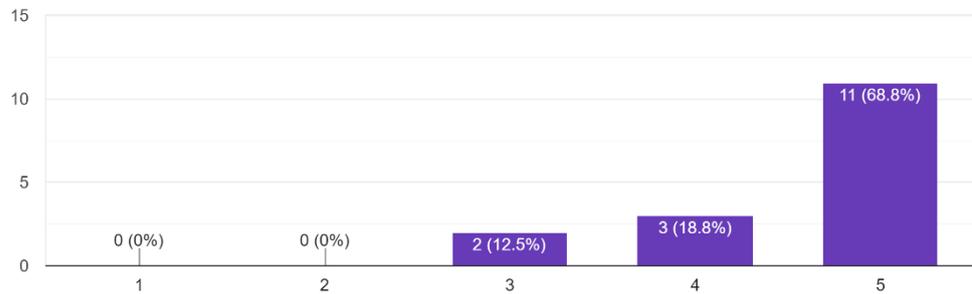


Gambar 5.10. Pertanyaan 2 Sample 2

Pilihan *Rating* adalah 1 (terlalu sedikit informasi) sampai 5 (terlalu banyak informasi) dengan skor paling baik ada 3 (cukup). Dari 16 responden, 75% memberi skor 3 dan 12,5% memberi skor 5.

- c. Apakah ringkasan yang dihasilkan relevan (tidak keluar dari konteks berita) dengan artikel berita?

Apakah ringkasan yang dihasilkan relevan (tidak keluar dari konteks berita) dengan artikel berita?
16 responses

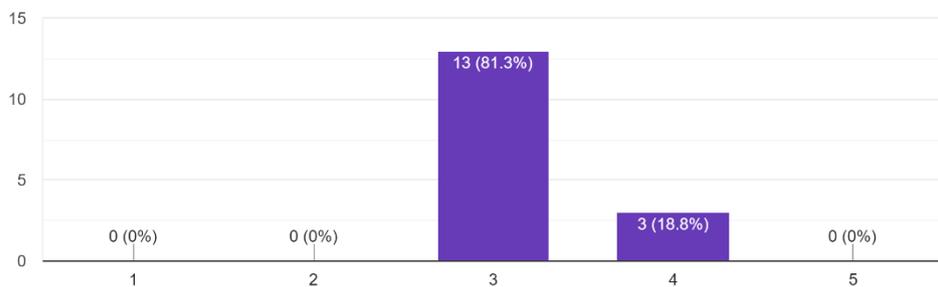


Gambar 5.11. Pertanyaan 3 Sample 2

Pilihan *Rating* adalah 1 (tidak relevan) sampai 5 (relevan). Dari 16 responden, 68,8% memberi skor 5 dan 18,8% memberi skor 4.

d. Apakah panjang ringkasan yang dihasilkan sudah cukup?

Apakah panjang ringkasan yang dihasilkan sudah cukup?
16 responses



Gambar 5.12. Pertanyaan 4 Sample 2

Pilihan *Rating* adalah 1 (terlalu pendek) sampai 5 (terlalu panjang) dengan skor terbaik adalah 3 (cukup). Dari 16 responden, 81,3% memberi skor 3 dan 18,8% memberi skor 4.

3. Sample 3

- **Artikel Berita**

pertumbuhan industri kecil dan menengah (ikm) terus digalakkan pemerintah untuk mendukung ekonomi nasional . berdasarkan laporan badan pusat statistik (bps)

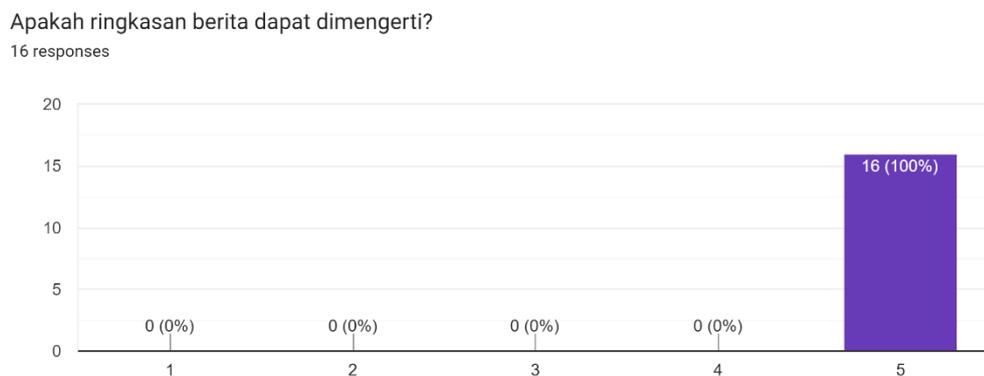
pertumbuhan industri tersebut terus meningkat dalam tiga tahun terakhir . pada 2014 , nilai tambah ikm tercatat mencapai rp __373__ triliun dan meningkat __17,6__ persen pada 2015 menjadi rp __439__ triliun . sementara sepanjang tahun lalu nilai tambahnya meningkat sekitar 18,3 persen . menteri keuangan sri mulyani mengatakan berbagai program dan kemudahan akan diberikan pemerintah pada __ukmk__ . selain pemberian kredit usaha rakyat (kur) dengan bunga rendah atau maksimal 9 persen , pemerintah juga sudah memberikan insentif pajak . " kalau terkait pajak , ukm itu pph (pajak penghasilan) , ppn (pajak penerimaan negara) nya berbeda berbeda dengan usaha yang lain . kalau dia omzet nya di bawah rp 4,8 miliar , pph nya final 1 persen dari peredaran bruto , " kata sri mulyani di kantor pusat astra international , sunter , jakarta utara , senin (3 / 4) . sementara jika ukm tersebut omzetnya sampai rp 50 miliar , akan mendapat fasilitas berupa pengurangan tarif sebesar 50 persen dari tarif normal , yang dikenakan dari penghasilan kena pajak dan bagian peredaran bruto sampai dengan rp 4,8 miliar . selain itu , pemerintah juga mencoba mengurangi beban ukm , yakni pengusaha kecil tidak wajib melaporkan usahanya untuk dikukuhkan sebagai pkp pengusaha kena pajak (pkp) dan tidak wajib memungut , menyetor , dan melaporkan ppn dan ppnbm (pajak penjualan atas barang mewah) .

- **Ringkasan Berita**

berdasarkan laporan badan pusat statistik (bps) pertumbuhan industri kecil dan menengah (ikm) terus meningkat dalam tiga tahun terakhir. berdasarkan pantauan bps, nilai tambah industri tersebut tumbuh mencapai rp 373 triliun.

- **Pertanyaan**

a. Apakah ringkasan berita dapat dimengerti?

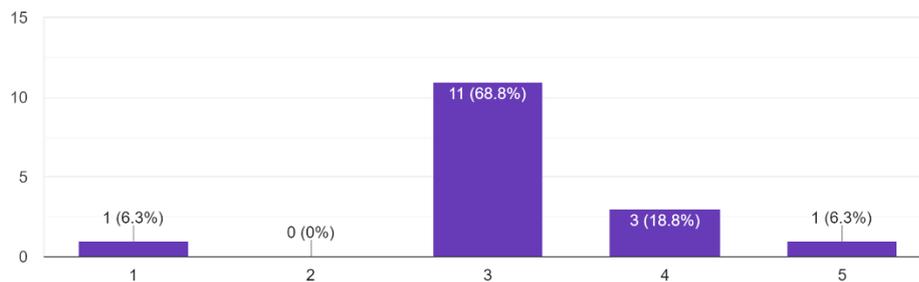


Gambar 5.13. *Pertanyaan 1 Sample 3*

Pilihan *Rating* adalah 1 (tidak dapat dimengerti) sampai 5 (dapat dimengerti). Dari 16 responden, 100% memberi skor 5.

- b. Apakah informasi yang dimasukkan ke dalam ringkasan sudah cukup? Note : Ringkasan yang baik mempertahankan inti sari dan menyingkirkan yang tidak penting.

Apakah informasi yang dimasukkan ke dalam ringkasan sudah cukup?
16 responses

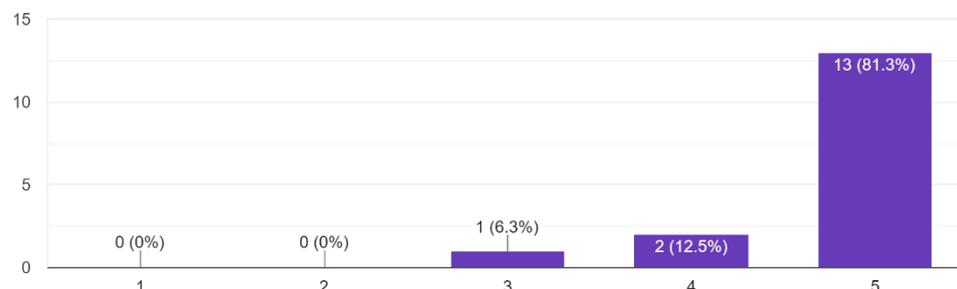


Gambar 5.14. Pertanyaan 2 Sample 3

Pilihan *Rating* adalah 1 (terlalu sedikit informasi) sampai 5 (terlalu banyak informasi) dengan skor paling baik ada 3 (cukup). Dari 16 responden, 68,8% memberi skor 3 dan 18,8% memberi skor 3.

- c. Apakah ringkasan yang dihasilkan relevan (tidak keluar dari konteks berita) dengan artikel berita?

Apakah ringkasan yang dihasilkan relevan (tidak keluar dari konteks berita) dengan artikel berita?
16 responses



Gambar 5.15. Pertanyaan 3 Sample 3

Pilihan *Rating* adalah 1 (tidak relevan) sampai 5 (relevan). Dari 16 responden, 81,3% memberi skor 5 dan 12,5% memberi skor 4.

d. Apakah panjang ringkasan yang dihasilkan sudah cukup?



Gambar 5.16. Pertanyaan 4 Sample 3

Pilihan *Rating* adalah 1 (terlalu pendek) sampai 5 (terlalu panjang) dengan skor terbaik adalah 3 (cukup). Dari 16 responden, 75% memberi skor 3 dan dan 128,8% memberi skor 4.

5.9. Kesimpulan Hasil Pengujian

Setelah melakukan berbagai pengujian, pada subbab ini akan dirangkum beberapa kesimpulan yang didapat.

5.9.1. Hasil Penilaian Model dengan Metrik

Model *Pointer-Generator Networks + BertSum-Extractive-IndoBERT* mendapatkan skor metrik yang cukup baik. Nilai metrik ROUGE dan BERTScore dari hasil ringkasan model PGN + BERTSum-Extractive-IndoBERT yaitu ROUGE-1 adalah 53,15%, ROUGE-2 adalah 43,32%, ROUGE-L adalah 51,10%, dan BERTScore adalah 81,05%.

5.9.2. *Out-of-Context Error* Belum Berhasil Diperbaiki

Model *Pointer-Generator Networks + BertSum-Extractive-IndoBERT* belum dapat memperbaiki *out-of-context error*. Hasil ringkasan pada PGN yang terlalu ekstraktif, sedangkan

hasil ringkasan IndoBERT yang terdapat *generator* terkadang menyebabkan *semantical error* saat berusaha untuk melakukan *text generation* untuk penggabungan frasa.

5.9.3. Model PGN terlalu Ekstraktif

Pointer-Generator Networks seharusnya bersifat *hybrid* dimana model dapat mengerti kondisi untuk mengekstrak kata dari artikel dan kondisi untuk melakukan *text generation*. Namun, karena dataset memiliki *label* yang sangat *ekstraktif*, maka model PGN belajar untuk menyesuaikan dengan apa yang dianggap benar, yaitu ringkasan referensi. Dengan pembelajaran yang condong kearah ekstraktif, model yang dihasilkan juga akan menjadi ekstraktif (karena objektif proses *training* adalah memperkecil nilai *loss* yang didapat dengan membandingkan hasil ringkasan dengan ringkasan referensi).

Pada skripsi ini, model PGN yang dihasilkan sama sekali tidak melakukan *text generation*, model hanya mengekstrak 2 hingga 4 kalimat pertama dari artikel. Hal ini juga disebabkan karena ringkasan referensi pada dataset cenderung mengambil 2 hingga 4 kalimat pertama. Oleh karena itu, model dapat memperoleh skor ROUGE yang baik. Nilai ROUGE menjadi buruk jika pada kondisi tertentu kalimat utama atau inti sari dari artikel berada pada tengah – tengah artikel.

5.9.4. Hallucination pada PGN tidak Ditemukan

Hal ini merupakan efek dari model PGN yang terlalu ekstraktif. *Hallucination* dapat terjadi ketika model salah mengeluarkan sebuah kata hasil proses *text generation*. Kata yang salah tersebut dapat berdampak pada kata – kata berikutnya. Namun, model PGN yang dihasilkan dari proses *training* menjadi ekstrim ekstraktif dan tidak melakukan proses *text generation*. Pada kasus ini, tidak ditemukan *hallucination* pada hasil ringkasan PGN karena semua teks diambil dari artikel.

5.9.5. Optimisasi Ringkasan dengan Model IndoBERT

Hasil ringkasan PGN akan menjadi *input* dari model IndoBERT dengan harapan akan mengoreksi kesalahan pada hasil ringkasan PGN. Jika pada ringkasan PGN terdapat *syntactical error*, IndoBERT dapat mengoreksi kesalahan sintaks. *Error* seperti kata repetitif, kata yang hilang atau kata yang tidak perlu dapat dikoreksi agar dapat lebih mudah dibaca. Namun, jika kesalahan pada ringkasan PGN berupa *hallucination*, IndoBERT tidak dapat menghilangkan hal tersebut.

IndoBERT pada dasarnya merupakan jenis model summarisasi ekstraktif, tapi IndoBERT juga dapat melakukan *text generation*. Berdasarkan analisis yang dilakukan, IndoBERT dapat melakukan *text generation* pada kondisi tertentu. Kondisi pertama merupakan ketika pada teks input ada kata yang hilang atau membutuhkan kata atau frasa untuk *bridging* antar kalimat atau frasa supaya lebih mudah dibaca. Kondisi kedua adalah perubahan kata dengan kata yang memiliki konteks yang mirip.