

Peringkasan Artikel Berita Menggunakan Pendekatan Abstraktif Dengan Model Transformers

1st Ananda Affan Fattahila
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia

affanfattahila@student.telkomuniversity.ac.id

2nd Ade Romadhony
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia

aderomadhony@telkomuniversity.ac.id

3rd Said Al Faraby
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia

saidalfaraby@telkomuniversity.ac.id

Abstrak— Membaca artikel berita merupakan kebiasaan rutin bagi mayoritas orang, di mana biasanya digunakan sebagai bahan bacaan atau referensi untuk mendapatkan informasi terbaru yang sedang beredar. Salah satu bentuk atau cara mendapatkan informasi secara cepat dapat dipermudah dengan keberadaan ringkasan artikel berita. Peringkasan artikel berita juga dapat meminimalkan informasi yang berlebihan. Tujuan Tugas Akhir ini adalah menghasilkan ringkasan dengan kualitas yang lebih baik pada beberapa topik berita yang performansinya rendah karena keterbatasan data. Pendekatan peringkasan yang digunakan adalah metode abstraktif dengan memanfaatkan pre-trained model berbasis transformers. Pre-trained model yang digunakan adalah T5-small dan BART-BASE serta untuk meningkatkan performansi, diterapkan proses augmentasi data. Pengujian pada Tugas Akhir ini dilakukan pada dataset XLSum Multi News, yang berisi kategori *government & politic, health, economic, infrastructure development, crime* dan *technology industry* dengan *crime* dan *economic* mempunyai kualitas ringkasan paling rendah. Evaluasi terhadap hasil ringkasan dilakukan secara otomatis dengan metrik ROUGE dan penilaian secara subyektif dari responden. Hasil evaluasi menunjukkan bahwa dari skor ROUGE tidak terjadi peningkatan performansi, namun responden menyimpulkan bahwa dari sisi relevansi, koherensi, dan kesesuaian, ringkasan yang dihasilkan setelah dilakukan augmentasi data mempunyai kualitas yang lebih baik.

Kata Kunci — Berita, Artikel, Peringkasan, Abstraktif, Augmentasi Data, Transformers

I. PENDAHULUAN

A. Latar Belakang

Artikel berita merupakan kebutuhan bagi banyak orang, di mana biasanya digunakan sebagai bahan bacaan atau referensi untuk mendapatkan suatu informasi terbaru yang sedang beredar. Membaca sebuah artikel berita juga merupakan kebiasaan rutin untuk mayoritas orang. Salah satu bentuk atau cara mendapatkan informasi dengan cepat pada artikel berita adalah dengan cara membaca ringkasannya. Peringkasan artikel berita juga memiliki

tujuan untuk meminimalkan informasi yang berlebihan [1]. Dalam dunia Natural Language Preprocessing terdapat hal yang menarik untuk mendapatkan informasi dari sebuah teks panjang. Hal tersebut dapat digunakan untuk peringkasan artikel berita untuk mendapatkan informasi yang ringkas dari sebuah teks [2]. Pendekatan peringkasan teks dapat dibagi menjadi dua kelompok: peringkasan ekstraktif dan peringkasan abstraktif. Dalam pendekatan ekstraksi, untuk mendapatkan kalimat kunci dan objek tanpa merubah objek itu sendiri adalah dengan metode ekstraksi. Metode ini dilakukan dengan menggunakan frasa kunci atau ekstraksi kalimat ad-hoc dan menjaga kalimat tetap utuh [4]. Sedangkan pada pendekatan abstraktif melibatkan parafrasa kalimat dengan memperhatikan konteks setelah memahami bahasa yang digunakan [5]. Selain itu peringkasan abstraktif membuat teks lebih sederhana dengan hanya menyajikan informasi penting dengan metode abstraksi yang menitikberatkan pada pemahaman konteks yang utama dari sumber dokumen dan membuat ringkasan yang singkat dan jelas dalam bahasa yang mudah dipahami [3]. Peringkasan teks abstraksi dianggap lebih rumit dibandingkan dengan peringkasan teks ekstraktif karena melibatkan pembuatan kalimat baru dengan berbagai metode pengulangan atau penggunaan kata-kata baru [6]. Teks yang diringkas menggunakan pendekatan abstraksi diharapkan dalam bentuk yang koheren sehingga dapat berkontribusi pada keterbacaan dan akurasi tata bahasanya [7].

Salah satu metode state-of-the-art peringkasan abstraktif menggunakan model fine-tuned generation adalah dengan menggunakan metode berbasis transformer (pada sistem Summaformer) [8]. Fine-tuned generation merupakan proses menyesuaikan model bahasa yang sudah dilatih sebelumnya (pre-trained) untuk tugas atau domain tertentu dengan dataset yang lebih kecil dan sistem Summaformer menduduki peringkat pertama pada tantangan peringkasan [8]. Selain itu dalam penelitian lain untuk tugas peringkasan abstraktif menggunakan fine-tuned model T5 memiliki keefektifan yang lebih bagus daripada BART [9].

Selain itu untuk meningkatkan akurasi model peringkasan abstraktif perlu dilakukan augmentasi data karena biasanya sangat terbatas. Hal ini dapat menyebabkan model tidak mampu menangani variasi dalam bahasa dan konteks, sehingga menyebabkan akurasi yang rendah. Salah satu penelitian menunjukkan bahwa penggunaan metode

augmentasi data dapat memperbaiki hasil ringkasan menjadi lebih baik [13].

Pada Tugas Akhir ini mengeksplorasi metode peringkasan artikel berita yang dapat menghasilkan ringkasan yang lebih baik dan efektif untuk mempermudah proses mendapatkan informasi dengan dilakukan percobaan augmentasi data berdasarkan klasifikasi topik. Hal ini dimotivasi untuk meningkatkan performansi model dengan menambahkan dataset sesuai dengan topik klasifikasi yang ditentukan. Selain itu juga untuk membuat model yang dibangun agar lebih memahami artikel berita pada kategori yang telah ditentukan. Dataset yang digunakan pada Tugas Akhir ini adalah XLSum dan Multi News yang terdiri atas kategori, yaitu government politic, health, economic, infrastructure development, crime, dan technology industry. Pada setiap kategori memiliki kata kunci yang ditentukan untuk melakukan pemilahan data.

1. Topik dan Batasan

Dataset yang digunakan berasal dari dua penelitian terkait peringkasan artikel berita yaitu XLSum dataset dari BBC News dan Multi News dengan menggunakan kategori augmentasi: government politic, health, economic, infrastructure development, crime, dan technology industry.

2. Tujuan

Tujuan Tugas Akhir ini adalah untuk memodelkan peringkasan artikel berita dengan pendekatan abstraktif dengan pengaplikasian augmentasi data. Pendekatan peringkasan yang digunakan adalah metode abstraktif dengan memanfaatkan pre-trained model berbasis transformers.

3. Organisasi Tulisan

Pada jurnal ini berisi bagian abstrak, pendahuluan, studi terkait, sistem yang dibangun, evaluasi, kesimpulan.

II. KAJIAN TEORI

A. Studi Terkait

1. Abstractive Text Summarization

Peringkasan merupakan tantangan penting pemahaman bahasa alami. Tujuannya adalah untuk menghasilkan suatu teks yang merepresentasikan makna sesungguhnya dari teks aslinya [10]. Sistem dari peringkasan merupakan sebuah proses menerima sebuah masukan berbentuk teks, kemudian ada proses yang melakukan peringkasan dan menghasilkan sebuah output berupa teks yang lebih singkat daripada masukannya.

Peringkasan dalam suatu teks natural language preprocessing dapat menggunakan pendekatan ekstraktif atau abstraktif. Peringkasan abstraktif mencoba untuk menghasilkan ringkasan dari bawah ke atas, sehingga aspek-aspek yang mungkin tidak muncul sebagai bagian

dari aslinya [5]. Sederhananya Abstraktif mengambil intisari dari suatu teks untuk menciptakan kalimat baru. Ilustrasi dari penerapan peringkasan abstraktif secara input dan output dapat dilihat pada tabel 1.

Input Teks Ringkasan	Output Teks Ringkasan
<i>The Met Office has issued a yellow weather warning for wind covering Wales and England, starting from 21:00 GMT on Wednesday evening.</i>	<i>Winds could reach gale force in Wales with stormy weather set to hit the whole of the country this week.</i>
<i>power are both likely to be disrupted, with the warning to remain in place until 15:00 on Thursday. Gusts of 55mph (88kmh) are likely and could hit up to 70mph on coasts and hills, with heavy and blustery showers.</i>	

TABLE 1.

Ilustrasi input dan output ringkasan abstraktif

2. Augmentasi Data

Augmentasi data merupakan teknik yang banyak digunakan dalam tugas sequence-to-sequence [11]. Dalam penelitian sebelumnya terjadi frustrasi karena kinerja suatu model terhadap data latih yang kecil, sehingga dilakukan augmentasi data dengan melakukan penambahan data latih. Penambahan ini terinspirasi dari penelitian visi komputer, yang hasilnya membantu melatih model menjadi lebih baik dalam tugasnya. Dalam prosesnya, sebelum proses penambahan data dilakukan, perlu dilakukan beberapa teknik untuk melakukan pemilahan data terlebih dahulu. Teknik yang digunakan dalam melakukan augmentasi data yaitu penggantian sinonim, random insertion kata, random swap kata dan penghapusan kata acak [12].

Penelitian augmentasi data pada tugas peringkasan teks berbasis abstraktif atau generatif [13], memiliki tujuan untuk membuat model yang dibangun menjadi dapat disesuaikan dengan kebutuhan informasi. Pada penelitian tersebut dataset yang digunakan berasal dari penelitian dataset CNN/Daily Mail [14], yang kemudian dimodifikasi menjadi sebuah dataset baru yaitu QMDS dataset [13]. Penggunaan teknik query yang diterapkan pada judul teks untuk menambang data train, dapat menghasilkan performansi model yang lebih baik setelah dilakukan augmentasi data.

3. Peringkasan Artikel Berbasis Transformers

Dalam peringkasan abstrak, model mencoba menghasilkan ringkasan yang alih-alih mengekstraksi kalimat atau kata kunci. Dibandingkan dengan ringkasan ekstraktif, ini lebih menantang dan membutuhkan skema pemodelan bahasa yang kuat untuk mencapai hasil yang baik. Secara tradisional, teknik ringkasan abstrak telah

berfokus pada menghasilkan teks pendek seperti judul atau headlines. Tetapi baru-baru ini, ada upaya untuk membuat ringkasan yang lebih panjang yaitu metode lama bergantung pada aturan transduksi pohon dan pendekatan tata bahasa quasi-synchronous untuk ringkasan abstrak yang efektif [8]. Baru-baru ini, pendekatan peringkasan saraf telah ditemukan lebih efektif. Model bahasa perwakilan saraf yang efektif sangat penting untuk tugas pembuatan teks. Dengan terobosan terbaru dari Transformer-based seperti menggunakan BERT dan BART, memanfaatkan jenis model ini sangat penting untuk mendapatkan representasi tekstual yang baik dalam tugas untuk peringkasan abstraksi [8].

Penelitian sebelumnya dalam tugas peringkasan teks menggunakan model BART dan T5, di mana kedua model tersebut diuji untuk membuktikan bahwa T5 memiliki keefektifan yang lebih baik dibandingkan dengan BART. Dalam beberapa aplikasi NLP, termasuk peringkasan teks, model bahasa pra-pelatihan telah menunjukkan kemajuan yang luar biasa. BART dan T5 adalah dua model pra-pelatihan canggih yang telah memberikan hasil yang sangat baik untuk tugas peringkasan teks. Dalam penelitian tersebut, telah disesuaikan untuk model-model tersebut dengan bantuan pembelajaran transfer untuk menghasilkan ringkasan artikel berita. Hasil menunjukkan bahwa t5 memberikan hasil yang lebih baik daripada dua model lain yang digunakan untuk pengujian [9].

4. Evaluasi

Evaluasi menjadi hal yang harus dilakukan untuk mengetahui model atau metode yang digunakan memiliki performansi yang baik atau tidak. Dalam Text Summarization Evaluasi yang digunakan biasanya menggunakan ROUGE. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) adalah metrik yang digunakan dalam NLP untuk mengevaluasi peringkasan teks. Metrik ini membandingkan ringkasan yang dibuat oleh model dengan ringkasan referensi yang dibuat oleh manusia atau ringkasan tujuan yang harus dicapai oleh model.

$$ROUGE\ N\ recall = \frac{Num\ word\ matches}{Num\ Words\ in\ references} \quad (1)$$

$$ROUGE\ N\ precision = \frac{Num\ word\ matches}{Num\ Words\ in\ summary} \quad (2)$$

$$ROUGE\ N\ F1\ score = \frac{precision \cdot recall}{precision + recall} \quad (3)$$

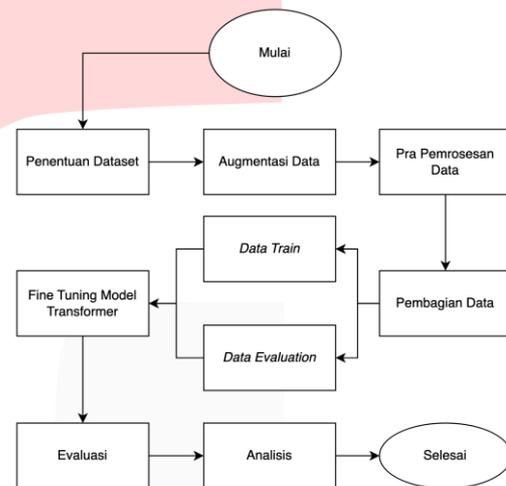
Berdasarkan formula diatas ROUGE-N mengukur tumpang tindih n-gram antara dua teks [15]. ROUGE-1 mengukur urutan pencocokan terpanjang dari kata-kata, tanpa panjang n-gram yang telah ditentukan, dan tidak memerlukan kecocokan berurutan. ROUGE-w mengukur urutan pencocokan terpanjang yang memperhitungkan pencocokan berurutan [16]. Versi penarikan ROUGE melaporkan rasio n-gram dalam referensi yang juga ada dalam ringkasan yang dihasilkan. Versi presisi ROUGE melaporkan rasio n-gram dalam ringkasan yang dihasilkan yang juga ada dalam ringkasan referensi. Versi skor F dari

ROUGE adalah rata-rata harmonik dari ROUGE yang presisi dan ROUGE yang dipanggil kembali.

III. METODE

A. Sistem yang Dibangun

Sistem yang dibangun pada tugas akhir ini adalah memanfaatkan model pre-trained untuk menghasilkan ringkasan abstrak yang lebih baik dengan cara melakukan perlakuan terhadap dataset XLSum dan Multi News. Data latih dari XLSum dilakukan klasifikasi dengan kata kunci untuk mewakili setiap kategori yang ditentukan. Data dengan kategori terendah dilakukan perlakuan augmentasi menggunakan data Multi News yang telah dikategorikan sama seperti data XLSum. Dalam tugas akhir ini model pre-trained yang dimanfaatkan yaitu T5-small dan BART-BASE. Dengan gambaran sistem umum seperti pada Gambar 1.



GAMBAR 1.

Alur Penelitian Peringkasan Artikel Berita

1. Pembangunan Data Latih

Sistem yang digunakan dalam pembangunan data latih baru dalam tugas peringkasan abstraktif artikel berita ini dilakukan augmentasi data terhadap dataset XLSum dengan menggunakan data Multi News. Dataset XLSum yang berasal dari BBC News dan dalam pembangunannya dilakukan oleh profesional bahasa dari seluruh dunia, selain itu dataset XLSum juga baik dalam sebuah pengaturan sumber daya rendah untuk menghemat komputasi [17]. Sedangkan Multi News merupakan data yang terdiri dari artikel berita dan ringkasan tulisan manusia dari situs newser.com [18]. Augmentasi data pada data latih diharapkan agar menambah performansi model untuk melakukan evaluasi terhadap data. Proses yang dilakukan dalam pembangunan data latih atau augmentasi data memiliki alur seperti pada gambar 2.

GAMBAR 2.



Alur pembangunan data latih

Berdasarkan Gambar 3 Dalam proses Klasifikasi Topik dilakukan eksplorasi data terlebih dahulu yang dilakukan pada data latih dengan cara melakukan klasifikasi berdasarkan kata kunci untuk setiap kategorinya. Eksplorasi ini digunakan untuk mengetahui jumlah data yang tersebar untuk masing-masing kategori dari teks berita pada data latih. Untuk topik kategori beserta setiap kata kunci yang penulis definisikan untuk digunakan dalam tugas akhir ini seperti pada tabel 2.

TABEL 2.
Topik dan kata kunci klasifikasi kategori

Topik	Kata Kunci
Govern ment & Politic	"government", "politic", "election", "political", "parties", "political party"
Health	"covid-19", "corona", "health", "covid", "vaccine", "medicine", "aids", "hiv", "virus"
Economi c	"economy", "economic", "recession", "inflation", "economical"
Infrastru cture & Develop ment	"infrastructure", "facilities", "train", "development", "construction"
Crime	"crime", "robbery", "theft", "sexual harassment", "illegal", "drug", "criminal", "law", "criminal"
Tech Industry	"tech", "technology", "industry", "startup", "start-up", "business"

Pada tabel 2 kategori serta kata kunci yang telah ditentukan digunakan untuk melakukan klasifikasi terhadap data latih dengan metode pencarian kata pada setiap teks. Setelah melakukan klasifikasi terhadap seluruh data latih, terdapat data yang bukan bagian dari topik *Government & Politic, Health, Economic, Infrastructure & Development, Crime, Tech Industry*. Data tersebut dijadikan sebagai kategori tersendiri atau kategori lain.

TABEL 3.
Topik dan kata kunci klasifikasi kategori

Topik	Jumlah Data
Government & Politic	66.470
Health	63.632
Economic	39.686
Infrastructure & Development	66.470
Crime	45.757
Tech Industry	79.278
Non Class	123.490

Pada data latih Multi News juga dilakukan eksplorasi data untuk kategori *Crime* dan *Economic* dikarenakan dua kategori ini menjadi fokus augmentasi data mempunyai kualitas ringkasan paling rendah berdasarkan jumlah data.

Berdasarkan hasil klasifikasi yang dilakukan dari jumlah data Multi News yang berjumlah total 44.972 data didapatkan *Economic* 6.934 data, dan *Crime* 15.222 data. Pada proses augmentasi data yang dilakukan adalah menambahkan data dengan topik *Crime* dan *Economic* dari data train Multi News ke data XLSum.

2. Pra-Pemrosesan Data

Dalam tahapan pra-pemrosesan data atau menyiapkan data sebelum diolah oleh model. Proses ini memiliki tujuan untuk mengubah data agar sesuai dengan kebutuhan model. Dalam tugas akhir ini pra-pemrosesan data dilakukan dengan menggunakan bantuan method AutoTokenizer dari transformers yang memungkinkan untuk membuat sebuah tokenizer yang telah dilatih sebelumnya dari repository model library yang dipilih pada tugas akhir ini yaitu T5-small dan BART-BASE. Tokenisasi menggunakan AutoTokenizer yang dilakukan untuk membuat suatu kata yang terpisah akan spasi menjadi satu token. Adapun contoh dari bentuk kalimat yang diubah menjadi token ada pada tabel 3.

TABEL 3.
Contoh tokenizer menggunakan AutoTokenizer.from_pretrained

Teks	Token	Pretrained Model
“berita ini sangat menarik sekali karena hari ini argentina juara piala dunia”	[36, 10694, 16, 23, 9965, 144, 1076, 9, 9629, 142, 4766, 23, 3, 4031, 35, 9, 3, 16250, 16, 23, 3, 9917, 77, 9, 3, 2047, 2551, 2816, 138, 9, 146, 29, 23, 9, 1]	T5-small
“berita ini sangat menarik sekali karena hari ini argentina juara piala dunia”	[0, 1943, 3119, 11, 118, 11944, 415, 604, 271, 967, 842, 330, 3644, 449, 1322, 2133, 1368, 1512, 11, 118, 29480, 1342, 1243, 14974, 1742, 181, 2617, 102, 30261, 493, 2]	BART-BASE

3. Peringkasan Artikel Berita

Peringkasan artikel berita pada tugas akhir ini menggunakan dua model pre-trained berbasis transformers yaitu T5-small dan BART-BASE. Model yang digunakan adalah model sequence-to-sequence language yang telah dilatih sebelumnya. Dalam tugas akhir ini model pre-trained dilakukan fine-tuning untuk menyelesaikan tugas peringkasan teks artikel berita, dengan parameter yang ditentukan seperti pada tabel 4.

Tabel 4. Hyperparameter

Parameter	Nilai	Model
Evaluation Strategy	epoch	T5-small, BART-BASE
Learning Rate	2e-5	T5-small, BART-BASE

Train Batch Size	16	T5-small,BART-BASE
Eval Batch Size	16	T5-small,BART-BASE
Weight Decay	0.01	T5-small,BART-BASE
Save Total limit	3	T5-small,BART-BASE
Number Train Epoch	10	T5-small,BART-BASE
Predict With Generate	true	T5-small,BART-BASE
Fp16	true	T5-small,BART-BASE

IV. HASIL DAN PEMBAHASAN

A. Evaluasi

1. Hasil Pengujian

Untuk menjalankan kode yang dibuat menggunakan framework PyTorch serta NVIDIA A100 40GB GPU dan Google Colab Pro dengan GPU Premium dengan memanfaatkan model pre-trained T5-small dan BART-BASE. Pada tugas akhir ini memiliki beberapa hasil pengujian otomatis dengan metrik ROUGE dan evaluasi manusia. Untuk melakukan evaluasi pada model T5-small dan BART-BASE menggunakan data testing dan dari XLSum. Di mana evaluasi menggunakan data testing dituliskan pada tabel 5.

TABEL 5.
Hasil Evaluasi Model

Model dan Perilaku	Train Loss	Val Loss	R1	R2	RL	RLSum	Gen Len
T5-small tanpa augmentasi	2.4494	2.2204	31.6534	10.0563	24.8104	24.8732	18.7913
T5-small dengan augmentasi	2.5563	2.2332	31.4802	9.9475	24.6687	24.7013	18.8025
BART-BASE tanpa augmentasi	1.5496	1.7506	38.5509	17.1804	31.6297	31.6993	19.6701
BART-BASE dengan augmentasi	1.6869	1.7450	38.4651	17.181	31.5511	31.6061	19.6752

Berdasarkan hasil evaluasi otomatis menggunakan ROUGE metrik pada tabel 5 model setelah dilakukan fine-tuning dapat dikatakan bahwa augmentasi terhadap data train XLSum tidak terlalu berpengaruh atau akurasi menurun namun tidak terlalu jauh, hal ini dapat dilihat dari hasil setiap ROUGE Evaluasi dari model T5-small dan BART-BASE. Selain evaluasi ROUGE dari hasil training menggunakan data latih dan data tes, Tugas akhir ini juga melakukan evaluasi ROUGE metrik untuk setiap kategori yang telah ditentukan yaitu crime, health, government & politic, economic, infrastructure development, technology & industry dan data yang tidak masuk kategori manapun pada tabel 6 dan tabel 7.

TABEL 6.
Hasil per kategori dengan menggunakan model T5-small

	Augmentasi Data (Tanpa Penambahan Data / Dengan Penambahan Data)				
	R1	R2	RL	RLSum	Gen Len
Crime	31.2004 / 31.0676	10.1739 / 9.9291	24.478 / 24.2629	24.577 / 24.3435	18.8553 / 18.8587
Health	29.8874 / 29.7895	9.1657 / 9.0008	23.3367 / 23.2296	23.4382 / 23.3414	18.8369 / 18.844
Government & Politic	30.1073 / 29.764	9.7306 / 9.3255	23.5127 / 23.3987	23.5644 / 23.4811	18.8177 / 18.8047
Economic	29.891 / 29.6269	9.6356 / 9.3641	23.4195 / 23.2046	23.4793 / 23.3023	18.7943 / 18.7943
Infrastructure Development	29.8882 / 29.8256	8.6619 / 8.7834	23.343 / 23.2681	23.4239 / 23.3099	18.8113 / 18.791
Technology & Industri	30.0241 / 29.8679	9.085 / 8.9511	23.3824 / 23.3813	23.4794 / 23.477	18.7977 / 18.799
Non Class	32.8254 / 32.6149	10.7059 / 10.4665	25.979 / 25.648	25.993 / 25.6484	18.8001 / 18.7863

TABEL 7.
Hasil per domain dengan menggunakan model BART-BASE

Kategori	Augmentasi Data (Tanpa Penambahan Data / Dengan Penambahan Data)				
	R1	R2	RL	RLSum	Gen Len
Crime	38.3095 / 37.9093	17.4337 / 16.976	31.4196 / 31.1729	31.5116 / 31.2852	19.7365 / 19.7078
Health	36.6147 / 36.3951	15.9566 / 15.6163	29.9192 / 29.6807	30.0282 / 29.7624	19.7052 / 19.6914
Government & Politic	37.4722 / 36.9109	16.9164 / 16.4513	30.4706 / 30.0311	30.5623 / 30.1347	19.6536 / 19.642
Economic	36.1042 / 35.9807	15.9539 / 15.7199	29.4693 / 29.688	29.7793 / 29.5647	19.6683 / 19.6192
Infrastructure Development	35.8679 / 36.1552	15.3094 / 15.3092	29.2172 / 29.4821	29.5538 / 29.302	19.68 / 18.7977
Technology & Industri	36.2867 / 36.0987	15.4874 / 15.366	29.5862 / 29.6086	29.6627 / 29.6894	19.6585 / 19.6074
Non Class	40.0295 / 40.0015	18.2205 / 18.2557	33.0573 / 33.1558	33.0823 / 33.1765	19.6653 / 19.6729

Dalam tugas akhir ini juga dilakukan *human evaluation* oleh 18 orang mahasiswa yang terdiri dari 17 mahasiswa tingkat akhir. Dalam proses *human evaluation* dinilai berdasarkan beberapa nilai untuk setiap ringkasan

yaitu Relevansi, Koherensi, dan Kesesuaian. Responden memberikan penilaian berupa *score* dengan rentang nilai 1-5, dimana nilai 5 menunjukkan sangat baik dan nilai 1 menunjukkan sangat buruk. Berikut merupakan hasil human evaluation dari model T5-small dan BART-BASE pada tabel 8, sedangkan pada tabel 9 merupakan nilai persentase kenaikan evaluasi setelah dilakukan augmentasi data.

TABEL 8.
Hasil Human Evaluation untuk skala 1 sampai 5

Model dan Perilaku	Rata-Rata Relevansi	Rata-Rata Koherensi	Rata-Rata Kesesuaian
T5-small tanpa augmentasi	3.0294	3.1353	2.8529
T5-small dengan augmentasi	3.3882	3.5118	3.2941
BART-BASE tanpa augmentasi	3.1353	3.2824	2.9588
BART-BASE dengan augmentasi	3.6352	3.7529	3.4941

2. Analisis Hasil Pengujian

Dalam analisis hasil pengujian menunjukkan bahwa penambahan dataset untuk transformer tidak selalu meningkatkan akurasi menggunakan metrik ROUGE, tetapi dapat meningkatkan akurasi menurut evaluasi manusia untuk relevansi T5-small 7% dan relevansi BART-BASE 10%, koherensi T5-small 8% dan koherensi BART-BASE 9%, dan kesesuaian T5-small 9% dan kesesuaian BART-BASE 11%. Relevansi mempresentasikan tingkat kecocokan antara teks yang dinilai dengan pertanyaan atau topik yang diberikan, koherensi mempresentasikan kemampuan teks untuk menjelaskan ide atau informasi secara jelas dan mudah dipahami, dan kesesuaian mempresentasikan tingkat keakuratan informasi yang disajikan dalam teks. Hal ini mungkin disebabkan bahwa ROUGE metrik yang lebih bersifat teknis dan hanya mengukur seberapa dekat kesamaan antara teks yang dihasilkan oleh model dengan teks yang dijadikan acuan. Sementara itu, evaluasi manusia dapat mengambil aspek-aspek lain seperti kegunaan dan kejelasan teks yang dihasilkan oleh model ke dalam pertimbangan. Oleh karena itu, meskipun akurasi ROUGE tidak meningkat, model mungkin masih menghasilkan teks yang lebih berkualitas menurut evaluasi manusia.

V. KESIMPULAN

A. Kesimpulan

Dalam tugas akhir ini dilakukan proses augmentasi data pada tugas peringkasan artikel berita. Penentuan augmentasi data dilakukan dengan klasifikasi data berdasarkan kata kunci dari dua dataset yaitu XLSum dan Multi News. Dalam proses fine-tuning model yang digunakan ada dua yaitu T5-small dan BART-BASE dengan parameter yang sama serta tempat pelatihan yang sama menggunakan PyTorch dan NVIDIA A100 40GB GPU dengan mendapatkan hasil training evaluasi yang menurun. Dalam evaluasi menggunakan data testing dari

XLSum didapatkan hampir semuanya tidak menunjukkan peningkatan nilai ROUGE setelah dilakukan augmentasi data. Sementara evaluasi manusia yang dilakukan terhadap kategori economic dan crime evaluasi manusia terkait relevansi, koherensi, dan kesesuaian meningkat untuk model yang menggunakan data yang sudah di augmentasi.

Berdasarkan hasil analisis pada tugas akhir ini, dapat disimpulkan bahwa penambahan dataset untuk transformer dapat meningkatkan akurasi menurut evaluasi manusia untuk relevansi, koherensi, dan kesesuaian, meskipun tidak selalu meningkatkan akurasi menggunakan metrik ROUGE. Oleh karena itu, saran yang dapat diberikan adalah untuk terus mengevaluasi hasil dari model menggunakan berbagai metrik yang berbeda, termasuk metrik teknis seperti ROUGE serta evaluasi manusia, untuk mendapatkan gambaran yang lebih luas tentang kualitas hasil yang dihasilkan oleh model. Selain itu, penting juga untuk memastikan bahwa dataset yang digunakan untuk pelatihan model representatif dari teks yang akan dikerjakan oleh model, karena ini dapat mempengaruhi kualitas hasil yang dihasilkan oleh model.

REFERENSI

- [1] El-Kassas, W.S., Salama, C.R., Rafea, A.A. and Mohamed, H.K., 2021. Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165, p.113679.
- [2] Yasunaga, M., Kasai, J., Zhang, R., Fabbri, A.R., Li, I., Friedman, D. and Radev, D.R., 2019, July. Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 7386-7393).
- [3] Janaki Raman, K. and Meenakshi, K., 2021. Automatic text summarization of article (NEWS) using lexical chains and wordnet—A review. *Artificial Intelligence Techniques for Advanced Computing Applications*, pp.271-282.
- [4] Badry, R.M., Eldin, A.S. and Elzanfally, D.S., 2013. Text summarization within the latent semantic analysis framework: Comparative study. *International Journal of Computer Applications*, 81(11), pp.40-45.
- [5] Rush, A.M., Chopra, S. and Weston, J., 2015. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*.
- [6] Ahmad, T., Ahmed, S.U., Ahmad, N., Aziz, A. and Mukul, L., 2020. News Article Summarization: Analysis and Experiments on Basic Extractive Algorithms. *International Journal of Grid and Distributed Computing*, 13(2), pp.2366-2379.
- [7] Moratanch, N. and Chitrakala, S., 2016, March. A survey on abstractive text summarization. In *2016 International Conference on Circuit, power and computing technologies (ICCPCT)* (pp. 1-7). IEEE.
- [8] Roy, S.G., Pinnaparaju, N., Jain, R., Gupta, M. and Varma, V., 2021. Summaformers@ LaySumm 20, LongSumm 20. *arXiv preprint arXiv:2101.03553*.
- [9] Garg, A., Adusumilli, S., Yenneti, S., Badal, T., Garg, D., Pandey, V., Nigam, A., Gupta, Y.K., Mittal, G. and Agarwal, R., 2021. NEWS article

- summarization with pretrained transformer. In International Advanced Computing Conference (pp. 203-211). Springer, Singapore.
- [10] Khatri, C., Singh, G. and Parikh, N., 2018. Abstractive and extractive text summarization using document context vector and recurrent neural networks. arXiv preprint arXiv:1807.08000.
- [11] Loem, M., Takase, S., Kaneko, M. and Okazaki, N., 2022. ExtraPhrase: Efficient Data Augmentation for Abstractive Summarization. arXiv preprint arXiv:2201.05313.
- [12] Wei, J. and Zou, K., 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. arXiv preprint arXiv:1901.11196.
- [13] Pasunuru, R., Celikyilmaz, A., Galley, M., Xiong, C., Zhang, Y., Bansal, M. and Gao, J., 2021, May. Data augmentation for abstractive query-focused multi-document summarization. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 15, pp. 13666-13674).
- [14] Hermann, K.M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M. and Blunsom, P., 2015. Teaching machines to read and comprehend. Advances in neural information processing systems, 28.
- [15] Kieuvongngam, V., Tan, B. and Niu, Y., 2020. Automatic text summarization of covid-19 medical research articles using bert and gpt-2. arXiv preprint arXiv:2006.01997.
- [16] ROUGE, L.C., 2004, July. A package for automatic evaluation of summaries. In Proceedings of Workshop on Text Summarization of ACL, Spain
- [17] Hasan, T., Bhattacharjee, A., Islam, M.S., Samin, K., Li, Y.F., Kang, Y.B., Rahman, M.S. and Shahriyar, R., 2021. XL-sum: Large-scale multilingual abstractive summarization for 44 languages. arXiv preprint arXiv:2106.13822.
- [18] Fabbri, A.R., Li, I., She, T., Li, S. and Radev, D.R., 2019. Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model. arXiv preprint arXiv:1906.01749.