

**PERINGKAS OTOMATIS BERITA BAHASA INDONESIA
PADA DOMAIN PEMILU MENGGUNAKAN BART**

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh:

**MUHAMMAD JAWAHIRUL FANANI
09020621036**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA
2025**

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini,

Nama : MUHAMMAD JAWAHIRUL FANANI
NIM : 09020621036
Program Studi : Sistem Informasi
Angkatan : 2021

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "PERINGKAS OTOMATIS BERITA BAHASA INDONESIA PADA DOMAIN PEMILU MENGGUNAKAN BART". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 2 Juni 2025
Yang Menyatakan,



Muhammad Jawahirul Fanani
NIM. 09020621036

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh,

Nama : MUHAMMAD JAWAHIRUL FANANI

NIM : 09020621036

Judul : PERINGKAS OTOMATIS BERITA BAHASA INDONESIA PADA
DOMAIN PEMILU MENGGUNAKAN BART

Ini telah diperiksa dan disetujui untuk diujikan

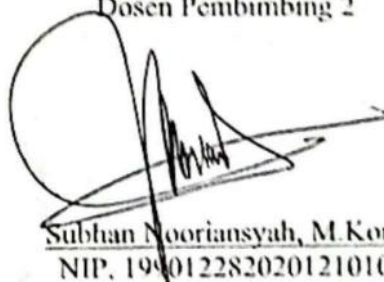
Surabaya, 28 Mei 2025

Dosen Pembimbing 1



Dr. Eng. Anang Kunaefi, M.Kom
NIP. 197911132014031001

Dosen Pembimbing 2




Subhan Nooriansyah, M.Kom
NIP. 199012282020121010

PENGESAHAN TIM PENGUJI SKRIPSI


Skripsi Muhammad Jawahirul Fanani ini telah dipertahankan
Di depan tim penguji Skripsi
Di Surabaya, 17 Juni 2025

Mengesahkan,
Dewan Penguji


Dosen Penguji 1


Khalid, M.Kom
NIP. 197906092014031002

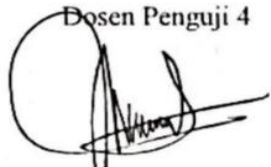
Dosen Penguji 2


Bayu Adhi Nugroho, Ph.D.
NIP. 197905182014031001

Dosen Penguji 3


Dr. Eng. Anang Kunaefi, M.Kom
197911132014031001

Dosen Penguji 4


Subhan Nooriansyah, M.Kom
NIP. 199012282020121010

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya


Dr. A. Saiful Hamdani, M.Pd
NIP. 196507312000031002



**KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Muhammad Jawahirul Fanani
NIM : 09020621036
Fakultas/Jurusan : Sains dan Teknologi/Sistem Informasi
E-mail address : jawahirulfn2002@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)
yang berjudul :

PERINGKAS TEKS OTOMATIS BERITA BAHASA INDONESIA

PADA DOMAIN PEMILU MENGGUNAKAN BART

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya dan menampilkan/mempublikasikannya di Internet atau media lain secara fulltext untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Sidoarjo, 8 Juli 2025

Penulis

(
Muhammad Jawahirul Fanani

ABSTRAK

PERINGKAS OTOMATIS BERITA BAHASA INDONESIA PADA DOMAIN PEMILU MENGGUNAKAN BART

By:

Muhammad Jawahirul Fanani

Tingginya volume berita online tentang pemilu selama masa pemilu di Indonesia, yang didorong oleh kemajuan teknologi, dapat menyulitkan publik dalam memahami informasi esensial terkait pemilu. Penelitian ini mengembangkan dan mengevaluasi model BART (*bart-base*) untuk meringkas berita pemilu berbahasa Indonesia, dengan fokus utama pada optimisasi *tokenizer*. Beberapa *custom tokenizer* berbasis *Byte-Pair Encoding* dikembangkan menggunakan korpus bahasa Indonesia. Selanjutnya, model BART di-*fine-tuning* dan dibandingkan dengan konfigurasi yang menggunakan *tokenizer* original sebagai *baseline*. Hasil menunjukkan bahwa semua model dengan *custom tokenizer* secara signifikan mengungguli *baseline*. Model '*bart-base-mix*' dan '*bart-base-detiknews-pemilu*' mendapatkan evaluasi terbaik, dengan skor tertinggi pada ROUGE-1 mencapai 25,40, ROUGE-2 mencapai 12,95, dan BERTScore mencapai 71,39. *Tokenizer* '*mix-50k*' terbukti paling efisien karena menghasilkan total token paling sedikit dan mencatatkan waktu pelatihan tercepat. Analisis stabilitas menunjukkan bahwa meskipun *baseline* memiliki standar deviasi terendah, model *custom* mampu mencapai stabilitas relatif (Koefisien Variasi) yang kompetitif, bahkan mengungguli *baseline* pada metrik ROUGE dengan nilai CV terendah mencapai 4,59% pada ROUGE-1 (*bart-base-detiknews-pemilu*) dan 8,07% pada ROUGE-2 (*bart-base-mix*). Pada metrik BERTScore *bart-base-baseline* memiliki nilai CV terendah yakni sebesar 0,34%.

Kata kunci: peringkas otomatis, transformer, BART, ROUGE, BERTScore

ABSTRACT

AUTOMATIC SUMMARIZATION OF INDONESIAN NEWS ON ELECTION DOMAIN USING BART

By:

Muhammad Jawahirul Fanani

The high volume of online news about elections during election periods in Indonesia, driven by technological advances, can make it difficult for the public to understand essential election-related information. This research develops and evaluates a BART (bart-base) model for summarizing Indonesian election news, with a particular focus on tokenizer optimization. Several custom tokenizers based on Byte-Pair Encoding are developed using Indonesian language corpus. Subsequently, the BART models were fine-tuned and compared with a configuration using the original tokenizer as the baseline. Results show that all models with the custom tokenizer significantly outperform the baseline. The models 'bart-base-mix' and 'bart-base-detiknews-pemilu' received the best evaluation, with the highest scores on ROUGE-1 reaching 25.40, ROUGE-2 reaching 12.95, and BERTScore reaching 71.39. The 'mix-50k' tokenizer proved to be the most efficient as it generated the fewest total tokens and recorded the fastest training time. The stability analysis shows that despite the baseline having the lowest standard deviation, the custom model is able to achieve competitive relative stability (Coefficient of Variation), even outperforming the baseline on the ROUGE metric with the lowest CV values reaching 4.59% on ROUGE-1 (bart-base-detiknews-pemilu) and 8.07% on ROUGE-2 (bart-base-mix). In the BERTScore metric bart-base-baseline has the lowest CV value of 0.34%.

Keywords: *auto summarizer, transformer, BART, ROUGE, BERTScore*

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	i
LEMBAR PENGESAHAN TIM PENGUJI SKRIPSI.....	ii
LEMBAR PERNYATAAN KEASLIAN SKRIPSI	iii
LEMBAR PERSETUJUAN PUBLIKASI	iv
MOTTO	v
KATA PENGANTAR.....	vi
ABSTRAK	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Penelitian	4
1.4 Tujuan penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Tinjauan Penelitian Terdahulu	6
2.2 Dasar Teori	8
2.2.1 Berita	8
2.2.2 Pemilu	9
2.2.3 Natural Language Processing (NLP)	10
2.2.4 Peringkat Teks Otomatis.....	11
2.2.5 Deep Learning.....	12
2.2.6 Transformer.....	13
2.2.7 Large Language Models.....	15
2.2.8 Transfer Learning.....	16
2.2.9 Bidirectional and Auto-Regressive Transformers (BART)	17
2.2.10 Byte Pair Encoding (BPE)	18

2.2.11	Metrik Fertility	22
2.2.12	ROUGE	22
2.2.13	BERTScore	24
2.2.14	Fleiss Kappa	25
2.2.15	Standar Deviasi	26
2.2.16	Dataset XL-Sum.....	27
2.3	Integrasi Keilmuan	28
BAB III METODE PENELITIAN		30
3.1	Alur Penelitian.....	30
3.1.1	Studi Literatur	31
3.1.2	Pengumpulan Data	31
3.1.3	Preprocessing Dataset <i>Fine-tuning</i>	33
3.1.4	Validasi Dataset <i>Fine-tuning</i>	34
3.1.5	Konfigurasi Tokenizer	34
3.1.6	Pemodelan BART	37
3.1.7	Evaluasi Model.....	38
3.1.8	Analisa Hasil	39
BAB IV HASIL & PEMBAHASAN		40
4.1	Pengumpulan Dataset	40
4.1.1	Dataset Fine-Tuning.....	40
4.1.2	Dataset Konfigurasi Tokenizer	41
4.2	Preprocessing Dataset <i>Fine-tuning</i>	42
4.2.1	Penghapusan Sumber Informasi.....	42
4.2.2	Penghapusan Nama dan Inisial Reporter	43
4.2.3	Penghapusan kalimat promo dan tautan.....	44
4.2.4	Penggabungan Dataset <i>Fine-tuning</i>	45
4.3	Validasi Dataset <i>Fine-tuning</i>.....	46
4.4	Konfigurasi Tokenizer	48
4.4.1	Preprocessing Korpus.....	48
4.4.2	Pelatihan Tokenizer.....	51
4.5	Pemodelan BART.....	63
4.5.1	Fine-tuning Skenario 1	65
4.5.2	Fine-tuning Skenario 2	67
4.5.3	Fine-tuning Skenario 3	69

4.5.3	Fine-tuning Skenario 4.....	71
4.6	Analisis Hasil	72
4.6.1	Analisis Tokenizer	73
4.6.2	Analisis Perbandingan Hasil Skenario Fine-tuning	76
4.6.3	Analisis Stabilitas Model	79
4.6.4	Hasil Ringkasan Otomatis Model	81
4.6.5	Perbandingan Model	83
BAB V	KESIMPULAN & SARAN	85
5.1	Kesimpulan	85
5.2	Saran Penelitian Berikutnya	86
DAFTAR PUSTAKA	87



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR GAMBAR

Gambar 2.1 Neural Network.....	12
Gambar 2.2 Arsitektur Transformers	13
Gambar 2.3 Blok Encoder Decoder BART.....	17
Gambar 2.4 Text corruption pre-training BART	18
Gambar 2.5 Flowchart pengisian vocabulary algoritma BPE.....	19
Gambar 2.6 Flowchart tokenisasi algoritma BPE	21
Gambar 3.1 Alur Penelitian.....	30
Gambar 4.1 Total Token Tokenizer indomixedtufs4.....	53
Gambar 4.2 Fertility Metric Tokenizer indmixedtufs4.....	54
Gambar 4.3 Total Token Tokenizer detiknews-pemilu	57
Gambar 4.4 Fertility Metric Tokenizer detiknews-pemilu	58
Gambar 4.5 Total Token Tokenizer mix.....	60
Gambar 4.6 Fertility Metric Tokenizer mix.....	61
Gambar 4.7 Total Token Setiap Tokenizer	74
Gambar 4.8 Fertility Metrics Setiap Tokenizer.....	75
Gambar 4.9 Perbandingan Rata-Rata Skor ROUGE Pada Set Tes.....	77
Gambar 4.10 Perbandingan Rata-Rata Skor BERTScore Pada Set Tes	78

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

Tabel 2.1 Daftar Penelitian Terdahulu.....	6
Tabel 2.2 Contoh n-gram.....	22
Tabel 2.3 Interpretasi Cohen Kappa.....	26
Tabel 3.1 Contoh dataset XI-Sum Indonesia.....	31
Tabel 3.2 Contoh dataset Liputan6.....	32
Tabel 3.3 Contoh dataset Indo-mixed-tufs4.....	33
Tabel 3.4 Contoh dataset detiknews-pemilu.....	33
Tabel 3.5 Skenario konfigurasi tokenizer.....	36
Tabel 3.6 Skenario eksperimen pelatihan BART.....	38
Tabel 4.1 Contoh dataset XI-Sum Pemilu.....	40
Tabel 4.2 Contoh dataset Liputan6 Pemilu.....	41
Tabel 4.3 Sampel Dataset Indo-mixed-tufs4.....	42
Tabel 4.4 Sampel Dataset detikNews Pemilu.....	42
Tabel 4.5 Hasil Penghapusan Sumber Informasi.....	43
Tabel 4.6 Hasil Penghapusan Nama dan Inisial Reporter.....	44
Tabel 4.7 Hasil Penghapusan Kalimat Promo dan Tautan.....	45
Tabel 4. 8 Detail identitas pakar.....	46
Tabel 4. 9 Hasil pelabelan pakar.....	46
Tabel 4.10 Interpretasi Tingkat Kesepakatan Fleiss Kappa.....	48
Tabel 4.11 Hasil Preprocessing Menghilangkan Pola Teks Spesifik.....	49
Tabel 4.12 Hasil Preprocessing Pemisahan Kata yang Gabung.....	50
Tabel 4.13 Hasil Preprocessing Pemisahan Kalimat.....	50
Tabel 4.14 Skenario Pelatihan Tokenizer.....	51
Tabel 4.15 Konfigurasi Tokenizer Skenario 1.....	53
Tabel 4.16 Hasil Konfigurasi Tokenizer Skenario 1.....	55
Tabel 4.17 Hasil tokenisasi variasi tokenizer indomixedtufs4.....	55
Tabel 4.18 Konfigurasi Tokenizer Skenario 2.....	56
Tabel 4.19 Hasil Konfigurasi Tokenizer Skenario 2.....	58
Tabel 4.20 Hasil tokenisasi variasi tokenizer detiknews-pemilu.....	59

Tabel 4.21 Konfigurasi Tokenizer Skenario 3	60
Tabel 4.22 Hasil Konfigurasi Tokenizer Skenario 3.....	62
Tabel 4.23 Hasil tokenisasi variasi tokenizer mix	62
Tabel 4.24 Proporsi dataset fine-tuning pada setiap fold.....	63
Tabel 4.25 Output BartTokenizer	64
Tabel 4.26 Skenario Fine-tuning BART	65
Tabel 4.27 Hyperparameter Fine-tuning Skenario 1.....	66
Tabel 4.28 Cross validation Fine-tuning skenario 1	66
Tabel 4.29 Rata-rata dan Standar Deviasi Sampel Fine-tuning Skenario 1.....	66
Tabel 4.30 Hyperparameter Fine-tuning Skenario 2.....	68
Tabel 4.31 Cross validation Fine-tuning skenario 2	68
Tabel 4.32 Rata-rata dan Standar Deviasi Sampel Fine-tuning Skenario 2.....	68
Tabel 4.33 Hyperparameter Fine-tuning Skenario 3.....	69
Tabel 4.34 Cross validation Fine-tuning skenario 3	70
Tabel 4.35 Rata-rata dan Standar Deviasi Sampel Fine-tuning Skenario 3.....	70
Tabel 4.36 Hyperparameter Fine-tuning Skenario 4.....	71
Tabel 4.37 Cross validation Fine-tuning skenario 4	71
Tabel 4.38 Rata-rata dan Standar Deviasi Sampel Fine-tuning Skenario 4.....	72
Tabel 4.39 Hasil Keseluruhan Skenario Konfigurasi Tokenizer	73
Tabel 4.40 Perbandingan Waktu Pelatihan Model.....	76
Tabel 4.41 Perbandingan Rata-Rata Set Test Skenario Fine-tuning.....	76
Tabel 4.42 Standar Deviasi Sampel Skenario Fine-tuning Set Tes	79
Tabel 4.43 Koefisien Variasi Skenario Fine-tuning Set Tes.....	80
Tabel 4.44 Hasil Ringkasan Setiap Model.....	82
Tabel 4.45 Perbandingan Model Dengan Penelitian Sebelumnya.....	83

DAFTAR PUSTAKA

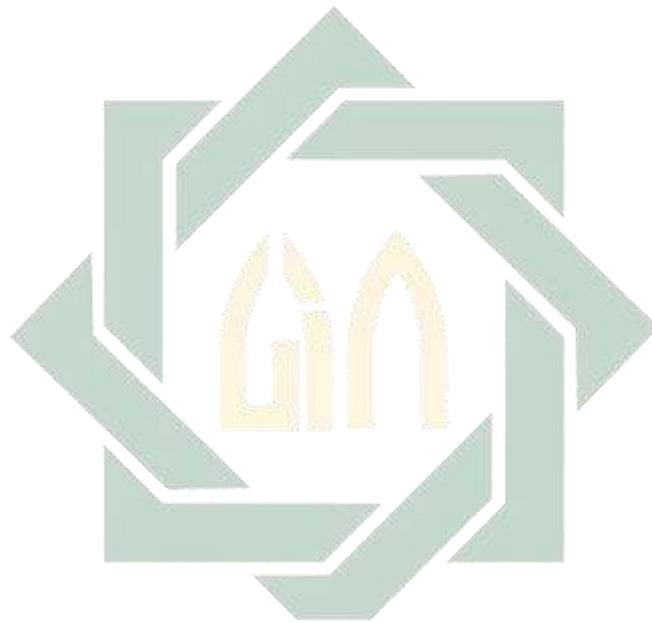
- Alami, N., Meknassi, M., & En-nahnahi, N. (2019). Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning. *Expert Systems with Applications*, 123, 195–211. <https://doi.org/10.1016/j.eswa.2019.01.037>
- Alyafeai, Z., Saeed AlShaibani, M., & Ahmad, I. (2020). *A SURVEY ON TRANSFER LEARNING IN NATURAL LANGUAGE PROCESSING A PREPRINT*.
- Astuti, R. H., Muljono, M., & Sutriawan, S. (2024). Indonesian News Text Summarization Using MBART Algorithm. *Scientific Journal of Informatics*, 11(1), 155–164. <https://doi.org/10.15294/sji.v11i1.49224>
- Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180–191. <https://doi.org/10.1177/0165551508095781>
- Cahyawijaya, S., Winata, G. I., Wilie, B., Vincentio, K., Li, X., Kuncoro, A., Ruder, S., Lim, Z. Y., Bahar, S., Khodra, M. L., Purwarianti, A., & Fung, P. (2021). *IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation*. <http://arxiv.org/abs/2104.08200>
- Deng, L., & Yu, D. (2013). Deep learning: Methods and applications. In *Foundations and Trends in Signal Processing* (Vol. 7, Issues 3–4, pp. 197–387). Now Publishers Inc. <https://doi.org/10.1561/20000000039>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <http://arxiv.org/abs/1810.04805>
- El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. In *Expert Systems with Applications* (Vol. 165). Elsevier Ltd. <https://doi.org/10.1016/j.eswa.2020.113679>
- Gage, P. (1994). A new algorithm for data compression. *The C Users Journal Archive*, 12, 23–38. <https://api.semanticscholar.org/CorpusID:59804030>
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). *Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages*. <http://corpora.uni-leipzig.de>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*.

- Gupta, A., Chugh, D., & Katarya, R. (2021). *Automated News Summarization Using Transformers*.
- Harahap, A. S. (2014). DAMPAK BERITA KRIMINAL DI TV. In *Dampak Berita Kriminal di TV Jurnal Komunikologi* (Vol. 11, Issue 2).
- Hasan, T., Bhattacharjee, A., Islam, M. S., Samin, K., Li, Y.-F., Kang, Y.-B., Rahman, M. S., & Shahriyar, R. (2021). *XL-Sum: Large-Scale Multilingual Abstractive Summarization for 44 Languages*. <http://arxiv.org/abs/2106.13822>
- Hikmat, Dr. H. M. M. (2018). *JURNALISTIK Literaly Journalism*.
- Jamanti, R. (2014). *PENGARUH BERITA BANJIR DI KORAN KALTIM TERHADAP KESADARAN LINGKUNGAN MASYARAKAT KELURAHAN TEMINDUNG PERMAI SAMARINDA*. 2(1), 17–33.
- Joshi, A., Fidalgo, E., Alegre, E., & Fernández-Robles, L. (2019). SummCoder: An unsupervised framework for extractive text summarization based on deep auto-encoders. *Expert Systems with Applications*, 129, 200–215. <https://doi.org/10.1016/j.eswa.2019.03.045>
- Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models Third Edition draft Summary of Contents*.
- Kamal Eddine, M., Tomeh, N., Habash, N., Le Roux, J., Vazirgiannis, M., & Polytechnique, É. (2022). *AraBART: a Pretrained Arabic Sequence-to-Sequence Model for Abstractive Summarization*. <https://huggingface.co/moussaKam/AraBART>
- Khan, B., Shah, Z. A., Usman, M., Khan, I., & Niazi, B. (2023). Exploring the Landscape of Automatic Text Summarization: A Comprehensive Survey. *IEEE Access*, 11, 109819–109840. <https://doi.org/10.1109/ACCESS.2023.3322188>
- Koto Fajri, Lau Jey Han, & Timothy Baldwin. (2020). *Liputan6: A Large-scale Indonesian Dataset for Text Summarization*. <https://github.com/>
- Kumar, Y. J., Goh, O. S., Basiron, H., Choon, N. H., & Suppiah, P. C. (2016). A review on automatic text summarization approaches. In *Journal of Computer Science* (Vol. 12, Issue 4, pp. 178–190). Science Publications. <https://doi.org/10.3844/jcssp.2016.178.190>
- La Quatra, M., & Cagliero, L. (2023). BART-IT: An Efficient Sequence-to-Sequence Model for Italian Text Summarization. *Future Internet*, 15(1). <https://doi.org/10.3390/fi15010015>

- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). *BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension*. <http://arxiv.org/abs/1910.13461>
- Lin, C.-Y. (2004). *ROUGE: A Package for Automatic Evaluation of Summaries*.
- Malte, A., & Ratadiya, P. (2019). *Evolution of Transfer Learning in Natural Language Processing*.
- Moratanch, N., & Chitrakala, S. (2017). *International Conference on Computer, Communication and Signal Processing : ICCSP 2017 : January 10-11, 2017, SSN College of Engineering, Chennai, Tamilnadu, India : special focus on IoT*. IEEE.
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). *A Comprehensive Overview of Large Language Models*. <http://arxiv.org/abs/2307.06435>
- Ni Luh Ratih Maha Rani. (2013). *Persepsi Jurnalis dan Praktisi Humas terhadap Nilai Berita*. 10.
- Ochin Sharma. (2019). *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing : trends, perspectives and prospects : COMITCON-2019 : 14th-16th February, 2019*. [IEEE].
- Parker, B., Sokolov, A., Ahmed, M., Kalebic, M., Kocak, S. A., & Shai, O. (2022). *Domain Specific Fine-tuning of Denoising Sequence-to-Sequence Models for Natural Language Summarization*. www.aaii.org
- Primakara, U. (2023). *IMPLEMENTASI PERINGKAS DOKUMEN BERBAHASA INDONESIA MENGGUNAKAN METODE TEXT TO TEXT TRANSFER TRANSFORMER (T5) I Nyoman Purnama 1) , Ni Nengah Widya Utami 2) Program Studi Sistem Informasi 1) , Sistem Informasi Akutansi 2)*.
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). *Pre-trained Models for Natural Language Processing: A Survey*. <https://doi.org/10.1007/s11431-020-1647-3>
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). *Deep Learning for Health Informatics*. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4–21. <https://doi.org/10.1109/JBHI.2016.2636665>
- Ruder, S., Peters, M., Swayamdipta, S., & Wolf, T. (2019). *Transfer Learning in Natural Language Processing Tutorial*.

- Rust, P., Pfeiffer, J., Vuli'cvuli'c, I., Ruder, S., & Gurevych, I. (2021). *How Good is Your Tokenizer? On the Monolingual Performance of Multilingual Language Models*. www.ukp.tu-darmstadt.de
- Samsudin, S. H., Dian Askhabul Yamin, Dr. Bambang Wahyu, Ferry Buchori Muslim, Dede Juhendi, & Deni Sediawan. (2020). *Buku Pintar Pemilu dan Demokrasi*.
- Schmitt, J. B., Debbelt, C. A., & Schneider, F. M. (2018). Too much information? Predictors of information overload in the context of online news exposure. *Information Communication and Society*, 21(8), 1151–1167. <https://doi.org/10.1080/1369118X.2017.1305427>
- Sennrich, R., Haddow, B., & Birch, A. (2016). *Neural Machine Translation of Rare Words with Subword Units*.
- Septriawati, D., Umboro Sumbar, D., & Bina Darma, U. (2024). IJM: Indonesian Journal of Multidisciplinary Pengaruh Konten Berita Terhadap Persepsi Politik. *IJM: Indonesian Journal of Multidisciplinary*, 2. <https://journal.csspublishing/index.php/ijm>
- Sydenham, P. H., & Thorn, R. (2005). *Handbook of measuring system design*. Wiley.
- Tao, C., Liu, Q., Dou, L., Muennighoff, N., Wan, Z., Luo, P., Lin, M., & Wong, N. (2024). *Scaling Laws with Vocabulary: Larger Models Deserve Larger Vocabularies*. <http://arxiv.org/abs/2407.13623>
- Tseng, B.-H., Cheng, J., Fang, Y., & Vandyke, D. (2020). *A Generative Model for Joint Natural Language Understanding and Generation*. Association for Computational Linguistics.
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*.
- We Are Social & Meltwater. (2024). *Digital 2024 Indonesia*. <https://datareportal.com/reports/digital-2024-indonesia>
- Wijaya, A. S., & Girsang, A. S. (2023). Augmented-Based Indonesian Abstractive Text Summarization using Pre-Trained Model mT5. *International Journal of Engineering Trends and Technology*, 71(11), 190–200. <https://doi.org/10.14445/22315381/IJETT-V71I11P220>
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). *BERTScore: Evaluating Text Generation with BERT*. <http://arxiv.org/abs/1904.09675>

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q.
(2019). *A Comprehensive Survey on Transfer Learning*.
<http://arxiv.org/abs/1911.02685>



UIN SUNAN AMPEL
S U R A B A Y A