



Evaluasi Analisa Sentimen pada Sistem Peringkasan Teks Otomatis menggunakan Model IndoBERT *Fine-Tuning* dalam Berita Berbahasa Indonesia

DOSEN PEMBIMBING

Dr. Surya Sumpeno, S.T., M.Sc.

Dr. Ir. Adhi Dharma Wibawa, S.T., M.T.

Program Studi Teknik Elektro Bidang Keahlian Telematika
Institut Teknologi Sepuluh Nopember



1. Pendahuluan

Dalam bab ini akan dijelaskan mengenai latar belakang masalah penelitian, penelitian terdahulu secara singkat, metode pemecahan masalah yang diusulkan, dan kebaruan atau kontribusi dari penelitian ini.

2. Penelitian Terkait

Pada bab ini, penelitian terdahulu akan dijelaskan secara lebih rinci yang akan menekankan kebaruan dari penelitian ini.

3. Metodologi

Desain sistem akan diperkenalkan dalam bab ini dan akan dijelaskan secara rinci.

OUTLINE

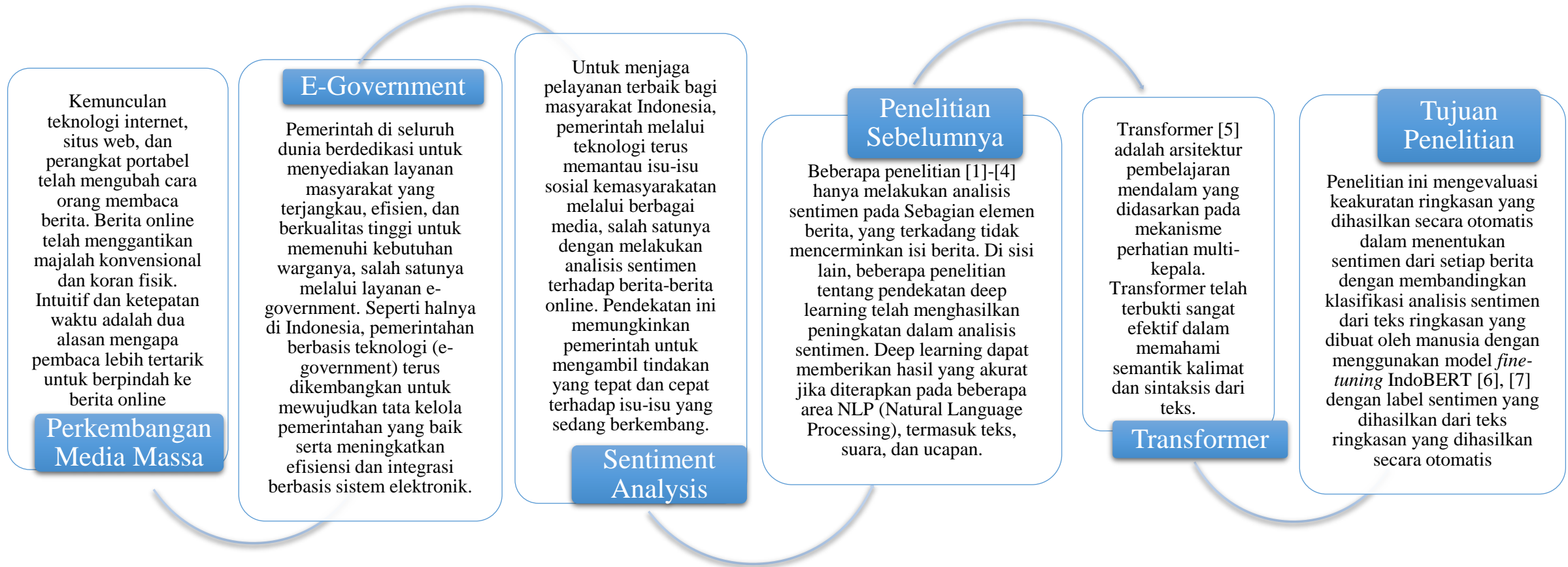
4. Hasil dan Diskusi

Hasil dari beberapa skenario pengujian akan dijelaskan. Data yang dihasilkan akan diinterpretasikan dan dijelaskan.

5. Kesimpulan

Poin-poin utama dari hasil penelitian akan dijelaskan dan saran-saran mengenai bagaimana penelitian ini dapat diperluas atau ditingkatkan juga akan diberikan

PENDAHULUAN



NEWS HEADLINE AND NEWS SUMMARY SENTIMENT COMPARISON



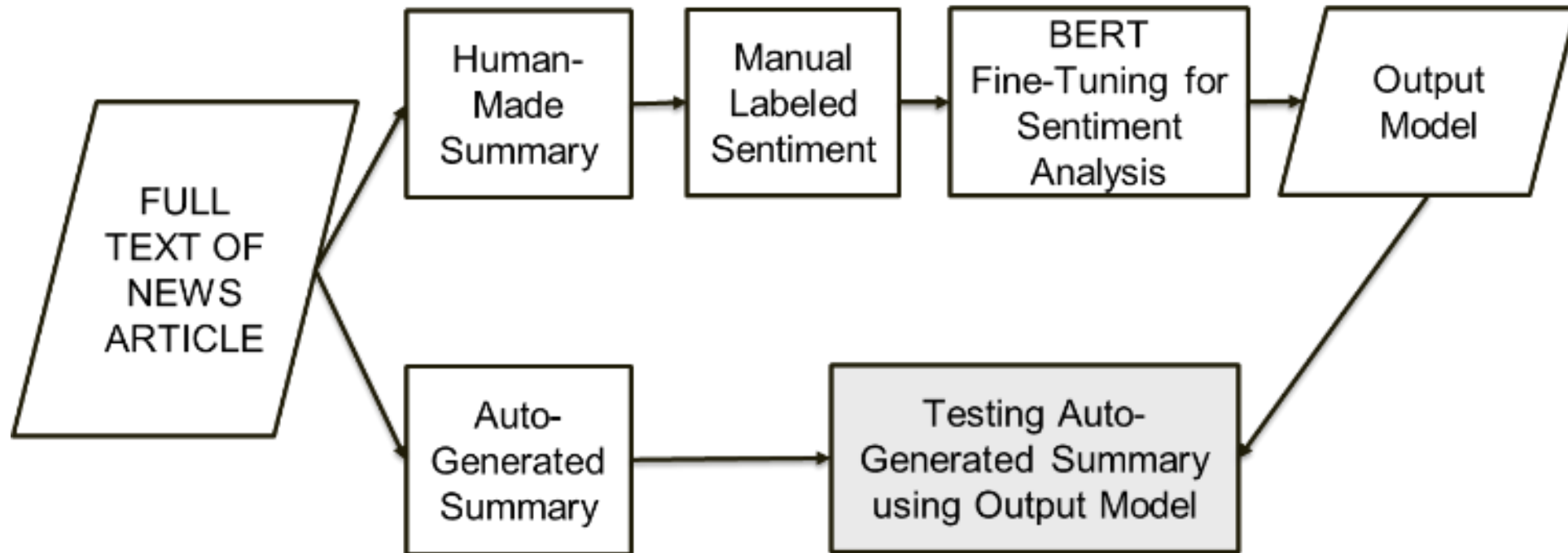
| | News Headline | Human-Made News Summary |
|----------|--------------------------------------|---|
| Content | Enrique: Neymar Seperti Penari Balet | “Pelatih Barcelona, Luis Enrique puas dengan keberhasilan timnya menghancurkan Villarreal dengan skor 4 - 1. Enrique pun tak lupa memuji Neymar yang disebutnya mirip penari balet di lapangan hijau. Neymar jadi pencetak gol pembuka bagi Barcelona di laga itu. Tak hanya itu, Neymar pun terus memberikan ancaman pada Villarreal sepanjang laga. Aksi Neymar mengolah bola mengundang decak kagum banyak orang, termasuk Enrique.” |
| Sentimen | Sentimen Negatif atau Positif? | Positif |

News source: CNN Indonesia web portal [8]

RELATED WORK

| No | Judul | Tahun | Teknik/Algoritma | Dataset | Hasil |
|----|--|-------|---|--|---|
| 1 | A Robustly Optimized BERT using Random Oversampling for Analyzing Imbalanced Stock News Sentiment Data | 2023 | Robustly optimized Bidirectional Transformer Encoder Representation (RoBERTa) | Stock News Dataset: 5.764 data(2.106 negative, 3.685 positive) | 84% F1-score, and 86% for accuracy |
| 2 | BERT Implementation on News Sentiment Analysis and Analysis Benefits on Branding | 2022 | BERTBASE-multilingual-cased | 1 600 News Titles | 93% accuracy |
| 3 | Sentiment Analysis of Economic News in Bahasa Indonesia Using Majority Vote Classifier | 2016 | Majority Vote Classifier | 540 News | Majority Vote scoring better precision and accuracy than decision tree, random forests, and support vector machine alone. |
| 4 | BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews | 2021 | IndoBERT-Base | 10.615 Google Play Indonesian User Review | akurasi terbaik: 84 |
| 5 | IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding | 2020 | IndoBERT versi IndoNLU | 12.760 data (online comment and review) | F1 Score: 92.72 |
| 6 | IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP | 2020 | IndoBERT versi IndoLEM | 5.048 sentiment data from Twitter and Hotel Review | F1 Score: 84.1 |

METHODOLOGY



Gambar 1. Desain Sistem

METHODOLOGY - Dataset

Tabel 1. Contoh dataset

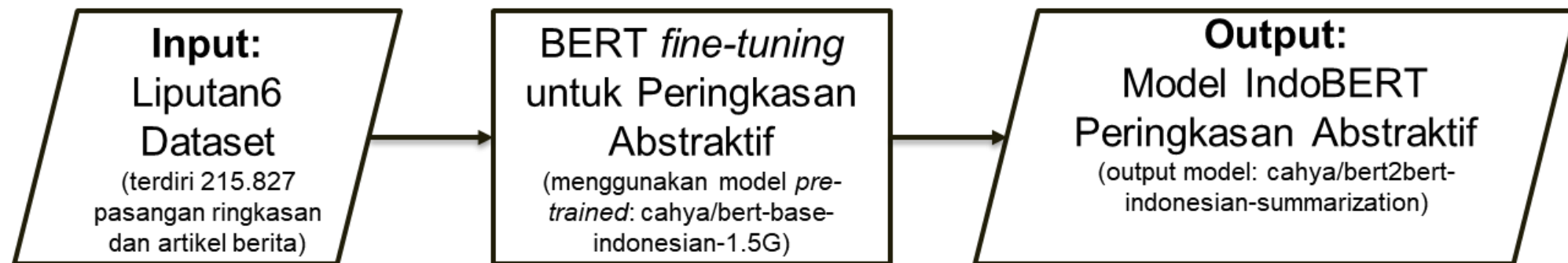
| Teks Berita | Ringkasan Berita Manual | Sentimen Manual |
|--|--|-----------------|
| Jakarta , CNN Indonesia - - Dokter Ryan Thamrin , yang terkenal lewat acara Dokter Oz Indonesia , meninggal dunia pada Jumat (4 / 8) dini hari . Dokter Lula Kamal yang merupakan selebriti sekaligus rekan kerja Ryan Thamrin menyebut kawannya itu sudah sakit sejak setahun yang lalu . Lula menuturkan , sakit itu membuat Ryan mesti vakum dari semua kegiatannya , termasuk menjadi pembawa acara Dokter Oz Indonesia . Kondisi itu membuat Ryan harus kembali ke kampung halamannya di Pekanbaru , Riau untuk menjalani istirahat . " Setahu saya dia orangnya sehat , tapi tahun lalu saya dengar dia sakit | Dokter Lula Kamal yang merupakan selebriti sekaligus rekan kerja Ryan Thamrin menyebut kawannya itu sudah sakit sejak setahun yang lalu . Lula menuturkan , sakit itu membuat Ryan mesti vakum dari semua kegiatannya , termasuk menjadi pembawa acara Dokter Oz Indonesia. ... | neutral |

Tabel 2. Komposisi Dataset

| Kategori Berita | negative | neutral | positive | Grand Total |
|--------------------|------------|------------|------------|-------------|
| hiburan | 7 | 48 | 30 | 85 |
| inspirasi | | 1 | 7 | 8 |
| olahraga | 50 | 82 | 116 | 248 |
| showbiz | 30 | 43 | 56 | 129 |
| tajuk utama | 124 | 162 | 110 | 396 |
| teknologi | 9 | 52 | 73 | 134 |
| Grand Total | 220 | 388 | 392 | 1000 |

METHODOLOGY - Model Peringkasan Otomatis

Gambar 2. Skema Fine-Tuning BERT untuk peringkasan Abstraktif



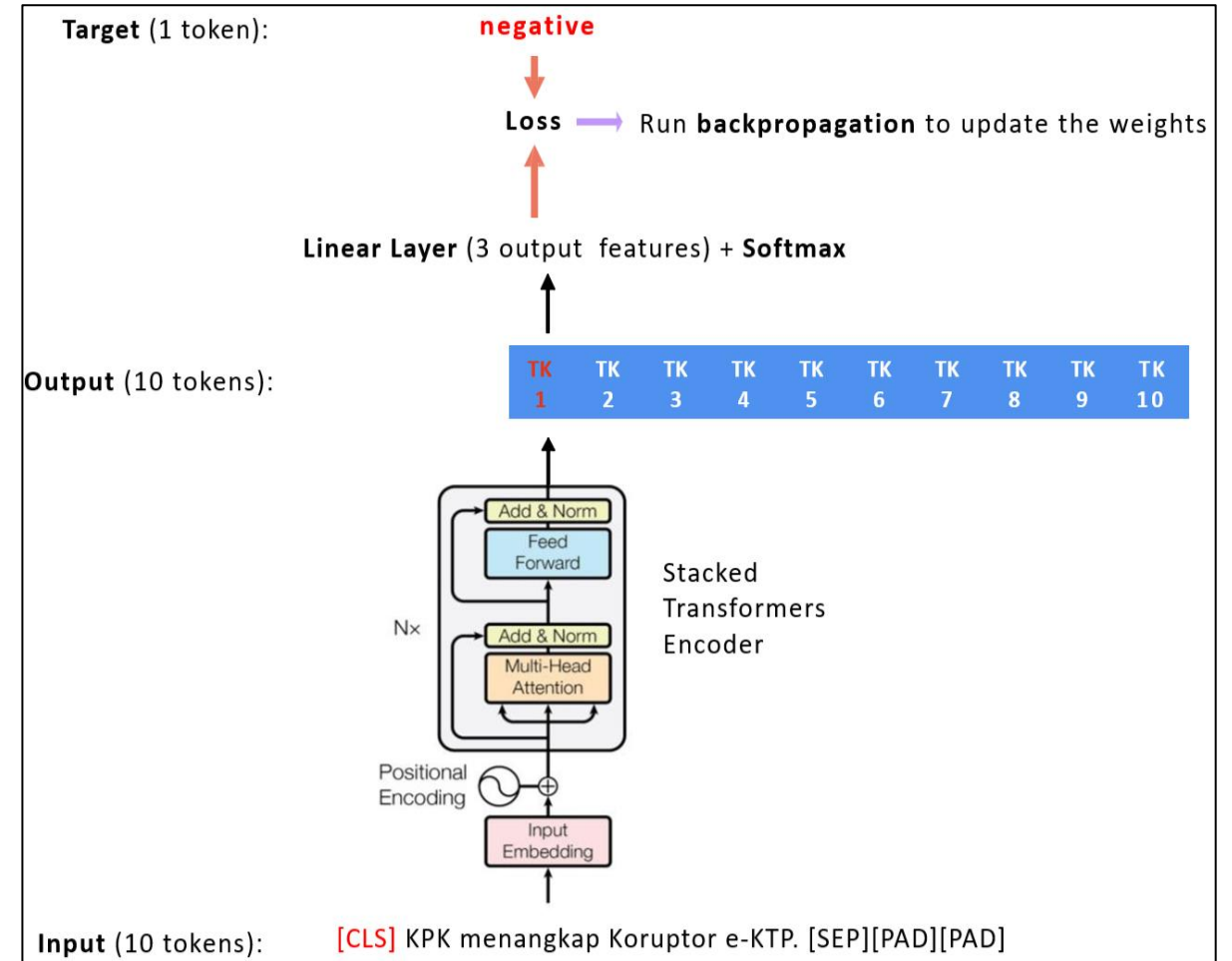
cahya/bert-base-indonesian-1.5G telah dilatih dengan 522MB Wikipedia bahasa Indonesia dan 1GB surat kabar berbahasa Indonesia yang berisi sekitar 500 ribu artikel (136 juta kata) dari 7 surat kabar Indonesia: Detik, Kompas, Tempo, CNN Indonesia, Sindo, Republika, dan Poskota

METHODOLOGY

Tabel 3. BERT Pre-training Model Details

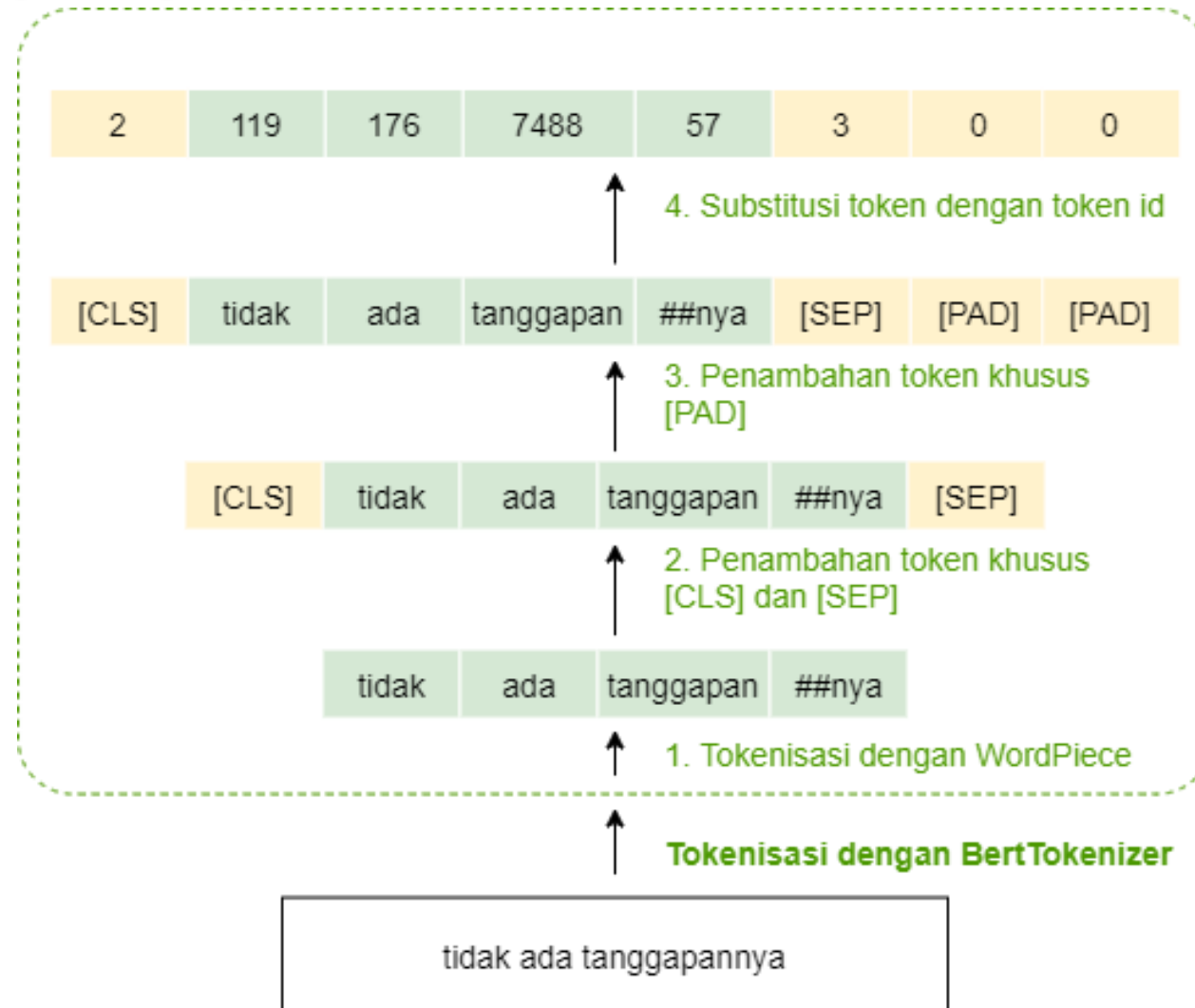
| Pre-Trained Model | L | H | A | Total Parameter | Language |
|------------------------|----|-----|----|-----------------|---------------------|
| BERT-base Multilingual | 12 | 768 | 12 | 110 million | 104 Language |
| IndoBERT-base IndoNLU | 12 | 768 | 12 | 124.5 million | Indonesian Language |
| IndoBERT-base IndoLEM | 12 | 768 | 12 | 110 million | Indonesian Language |

Dataset dibagi menjadi tiga bagian untuk tahap *fine-tuning*: pelatihan, pengujian, dan validasi. Set pelatihan memiliki 900 data, sedangkan set pengujian dan validasi masing-masing memiliki 50 data.



Gambar 3. The BERT Fine-Tuning for classification task

METHODOLOGY



Hasil Pengujian Fine-tuning IndoBERT menggunakan Ringkasan Berita Manual



Tabel 2. ML Baseline F1-Score

| ML Baseline Model | Training F1-Score | Cross Validation F1-Score | Testing F1-Score |
|-------------------|-------------------|---------------------------|------------------|
| SVM | 0.9967 | 0.4969 | 0.5921 |
| kNN | 0.6091 | 0.4086 | 0.4641 |
| Naïve Bayes | 0.9822 | 0.4863 | 0.5067 |
| Decision Tree | 1 | 0.4707 | 0.4227 |
| Random Forest | 1 | 0.4634 | 0.4244 |

Tabel 3. Pre-trained Models Performance with Learning Rate 1e-5, Trained and Tested with Human Summaries

| Pre-trained Model | Batch Size / Epochs | Avg Training F1-Score | Avg Validation F1-Score | Testing F1-Score |
|--------------------------------|---------------------|-----------------------|-------------------------|------------------|
| bert-base-multilingual-cased | 16 / 10 | 0.7933 | 0.6080 | 0.6124 |
| bert-base-multilingual-uncased | 16 / 10 | 0.7468 | 0.5428 | 0.5931 |
| indobenchmark/indobert-base-p1 | 16 / 10 | 0.8589 | 0.6733 | 0.7357 |
| indolem/indobert-base-uncased | 16 / 10 | 0.7827 | 0.6527 | 0.7214 |

Model yang sudah dilatih jauh mengungguli model pembelajaran mesin Baseline. Model pre-trained dengan performa tertinggi adalah indobenchmark/indobert-base-p1 dengan nilai F1-score sebesar 0.7357, kemudian di posisi kedua adalah indolem/indobert-base-uncased dengan nilai F1-score sebesar 0.7214.

IndoBERT juga mendapatkan skor yang lebih baik daripada versi multibahasa. Hasil ini terjadi karena kedua model IndoBERT ini telah dilatih sebelumnya, dan pelatihannya berfokus pada bahasa monolingual, sama dengan bahasa pada tugas target dalam bahasa Indonesia.

Hasil Pengujian Fine-tuning IndoBERT menggunakan Ringkasan Berita Manual



Tabel 4. Indonesian Pre-trained Models Performance With Learning Rate 1e-5, Trained And Tested With Human Summaries

| Pre-trained Model | Batch Size / Epochs | Avg Training F1-Score | Avg Validation F1-Score | Testing F1-Score |
|--------------------------------|---------------------|-----------------------|-------------------------|------------------|
| indobenchmark/indobert-base-pl | 16 / 5 | 0.7268 | 0.6607 | 0.6138 |
| indolem/indobert-base-uncased | 16 / 5 | 0.626 | 0.6513 | 0.7152 |
| indobenchmark/indobert-base-pl | 16 / 10 | 0.8589 | 0.6733 | 0.7357 |
| indolem/indobert-base-uncased | 16 / 10 | 0.7827 | 0.6527 | 0.7214 |
| indobenchmark/indobert-base-pl | 16 / 20 | 0.9294 | 0.6977 | 0.7111 |
| indolem/indobert-base-uncased | 16 / 20 | 0.8868 | 0.6507 | 0.7514 |

Seperti yang dapat kita lihat pada Tabel, skor tes yang dihasilkan oleh model IndoLEM meningkat seiring dengan bertambahnya epoch, yang mengindikasikan pembelajaran yang lebih baik dibandingkan dengan model IndoNLU. Hal ini terjadi karena, pada fase pra-pelatihan, model IndoLEM dilatih pada kumpulan teks yang berisi teks yang lebih formal, seperti Wikipedia dan teks berita, sesuai dengan teks berita dalam dataset penelitian. Kita akan menggunakan kedua versi Indobert ini untuk menguji dataset ringkasan berita yang dihasilkan secara otomatis.

Hasil Pengujian Fine-tuning IndoBERT menggunakan Ringkasan Berita Otomatis



Tabel 5. Pre-trained Models Performance With Learning RATE 1E-5, USE Auto-generated Summaries (AGS) Testing Set

| Pre-trained Model | Batch Size / Epochs | Avg Training F1-Score | Avg Validation F1-Score | AGS Testing F1-Score |
|--------------------------------|---------------------|-----------------------|-------------------------|----------------------|
| indobenchmark/indobert-base-pl | 16 / 10 | 0.8589 | 0.6733 | 0.5187 |
| indolem/indobert-base-uncased | 16 / 20 | 0.8868 | 0.6507 | 0.6541 |

Analisis sentimen menggunakan dataset pengujian ringkasan manusia memiliki nilai yang lebih baik dibandingkan dengan dataset pengujian ringkasan yang dihasilkan secara otomatis. Hasil ini terjadi karena ringkasan yang dibuat secara otomatis terkadang menghasilkan ringkasan yang hanya mengambil sebagian dari konten berita, sehingga tidak dapat menangkap semua informasi dan terkadang membuat kalimat yang tidak berhubungan.

| | |
|-------------------------------|--|
| HUMAN-MADE SUMMARY | “Angel Investment Network Indonesia (ANGIN) menginisiasi kegiatan Wonder Tech. Acara yang bertujuan untuk memahami permasalahan dan mencoba mencari solusi terbaik agar lebih banyak lagi perempuan terjun ke dunia teknologi, didukung oleh para pelaku startup. Dalam kegiatan tersebut akan dihadirkan beberapa tokoh perempuan yang sukses dengan bisnisnya di Indonesia untuk membahas potensi serta solusi terbaik untuk perempuan di dunia teknologi.” |
| AUTO-GENERATED SUMMARY | “permasalahan masih minimnya jumlah perempuan yang berkecimpung dalam dunia teknologi hingga saat ini ternyata mendapat sorotan dari para pelaku ukm hingga jajaran eksekutif di perusahaan teknologi di indonesia. namun menjamurnya jumlah responden dan meningkatnya lowongan posisi untuk engineer, ternyata tidak disertai dengan meningkatnya jumlah pelatihan yang berasal dari kalangan perempuan untuk berkarier sebagai manage untuk industri teknologi, tidaklah demikian dengan banyaknya wanita yang bekerja sebagai engineer.” |



“Menyebabkan Kesalahan Sentimen”

KESIMPULAN

Analisis sentimen pada ringkasan berita dapat dilakukan dengan menggunakan model BERT yang telah dilatih dalam bahasa Indonesia (IndoBERT). Kinerja prediksi analisis sentimen menggunakan model *fine-tuned* IndoBERT pada dataset ringkasan berita buatan manusia mengungguli model *machine learning* dan model *fine-tuned* BERT yang dilatih pada multibahasa. *Fine-tuning* menggunakan model *pre-trained indolem/indobert-base-uncased* menghasilkan nilai terbaik pada dataset pengujian ringkasan manusia, dengan skor F1 75%, dan menghasilkan nilai terbaik pada dataset pengujian ringkasan yang dibuat secara otomatis, dengan skor F1 65%. Hal ini menunjukkan bahwa kinerja prediksi analisis sentimen menggunakan peringkasan berita otomatis belum dapat melampaui kinerja prediksi analisis sentimen menggunakan ringkasan berita yang dibuat oleh manusia karena ringkasan yang dibuat secara otomatis terkadang menghasilkan ringkasan yang hanya mengambil sebagian dari isi berita dan terkadang membuat kalimat-kalimat yang tidak berkaitan yang tentunya berakibat pada penentuan sentimen berita yang salah.

SARAN

Untuk penelitian selanjutnya, memperbanyak dataset dan menggunakan model peringkasan teks otomatis yang menghasilkan ringkasan dengan alur informasi yang baik, dengan kalimat-kalimat yang berhubungan satu sama lain dan membentuk sebuah narasi yang dapat dimengerti oleh pembaca, layaknya ringkasan yang dibuat oleh manusia, sangat disarankan.

REFERENCES

- [1] S. M. Permataning Tyas, R. Sarno, A. T. Haryono, and K. Rossa Sungkono, “A Robustly Optimized BERT using Random Oversampling for Analyzing Imbalanced Stock News Sentiment Data,” ICCoSITE 2023 - International Conference on Computer Science, Information Technology and Engineering: Digital Transformation Strategy in Facing the VUCA and TUNA Era, pp. 897–902, 2023, doi: 10.1109/ICCoSITE57641.2023.10127725.
- [2] A. Hossain, M. Karimuzzaman, M. M. Hossain, and A. Rahman, “Text mining and sentiment analysis of newspaper headlines,” Information (Switzerland), vol. 12, no. 10, Oct. 2021, doi: 10.3390/info12100414.
- [3] M. F. Abdussalam, D. Richasdy, and M. Arif Bijaksana, “BERT Implementation on News Sentiment Analysis and Analysis Benefits on Branding,” JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 6, pp. 2064–2073, Oct. 2022, doi: 10.30865/mib.v6i4.4579.
- [4] Zamahsyari and A. Nurwidiantoro, “Sentiment analysis of economic news in Bahasa Indonesia using majority vote classifier,” in 2016 International Conference on Data and Software Engineering (ICoDSE), 2016, pp. 1–6. doi: 10.1109/ICoDSE.2016.7936123.
- [5] A. Vaswani et al., “Attention is All you Need,” in Advances in Neural Information Processing Systems, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- [6] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” Proceedings of the 28th International Conference on Computational Linguistics, pp. 757–770, Nov. 2020, [Online]. Available: <http://arxiv.org/abs/2011.00677>
- [7] B. Wilie et al., “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” Sep. 2020, Accessed: Sep. 21, 2022. [Online]. Available: <http://arxiv.org/abs/2009.05387>
- [8] “Enrique: Neymar Seperti Penari Balet.” Accessed: Dec. 15, 2023. [Online]. Available: <https://www.cnnindonesia.com/olahraga/20170507044007-142-212861/enrique-neymar-seperti-penari-balet/>
- [9] B. Liu, “Sentiment Analysis and Opinion Mining,” Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.
- [10] E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, “Sentiment Analysis Is a Big Suitcase,” IEEE Intell Syst, vol. 32, no. 6, pp. 74–80, 2017, doi: 10.1109/MIS.2017.4531228.
- [11] L. Zhang, S. Wang, and B. Liu, “Deep Learning for Sentiment Analysis : A Survey,” WIREs Data Mining and Knowledge Discovery, vol. 8, no. 4, p. e1253, Jan. 2018, Accessed: Dec. 03, 2022. [Online]. Available: <http://arxiv.org/abs/1801.07883>
- [12] Jason Brownlee, “A Gentle Introduction to Text Summarization.” Accessed: Dec. 02, 2022. [Online]. Available: <https://machinelearningmastery.com/gentle-introduction-text-summarization>
- [13] I. R. Musyaffanto, G. Budi Herwanto, and M. Riasetiawan, “Automatic extractive text summarization for indonesian news articles using maximal marginal relevance and non-negative matrix factorization,” Proceedings - 2019 5th International Conference on Science and Technology, ICST 2019, no. July, 2019, doi: 10.1109/ICST47872.2019.9166376.
- [14] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, vol. 1, pp. 4171–4186, Oct. 2018, doi: 10.48550/arxiv.1810.04805.

THANK
YOU

