

**ANALISIS DETEKSI EMOSI PADA TEKS BAHASA INDONESIA DAN
BAHASA JAWA DENGAN METODE CROSS-LINGUAL
MENGUNAKAN MODERN MULTILINGUAL BERT**

SKRIPSI



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

Disusun Oleh:

PUTRI RORO ANDINI

09020622037

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

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : Putri Roro Andini

NIM : 09020622037

Program Studi : Sistem Informasi

Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: **"ANALISIS DETEKSI EMOSI PADA TEKS BAHASA INDONESIA DAN BAHASA JAWA DENGAN METODE CROSS-LINGUAL MENGGUNAKAN MODERN MULTILINGUAL BERT"**. 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, 1 Januari 2026

Yang menyatakan



(Putri Roro Andini)

NIM 09020622037

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA : PUTRI RORO ANDINI
NIM : 09020622037
JUDUL : ANALISIS DETEKSI EMOSI PADA TEKS BAHASA
INDONESIA DAN JAWA DENGAN METODE
CROSS-LINGUAL MENGGUNAKAN MODERN
MULTILINGUAL BERT

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 29 Desember 2025

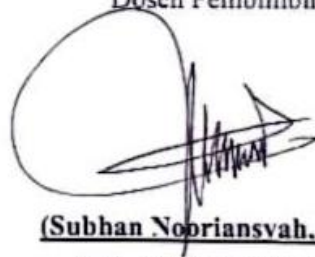
Dosen Pembimbing 1



(Dr. Eng. Anang Kunaefi, M.Kom)

NIP. 197911132014031001

Dosen Pembimbing 2



(Subhan Nooriansyah, M.Kom.)

NIP. 199012282020121010

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Putri Roro Andini ini telah dipertahankan
di depan tim penguji skripsi
di Surabaya, 2 Januari 2026

Mengesahkan,
Dewan Penguji

Penguji I

(Khalid, M.Kom.)

NIP. 197906092014031002

Penguji II

(Yunita Ardilla, S.Kom., M.MT.)

NIP. 199206042020122015

Penguji III

(Dr. Eng. Anang Kundaefi, M.Kom.)

NIP. 197911132014031001

Penguji IV

(Subhan Nooriansyah, M.Kom.)

NIP. 199012282020121010

Mengetahui,

Dekan Fakultas Sains dan Teknologi

UIN Sunan Ampel Surabaya



(Dr. Asep Saepul Hamdani, M.Pd.)

NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

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 : Putri Roro Andini
NIM : 09020622037
Fakultas/Jurusan : Sains dan Teknologi / Sistem Informaasi
E-mail address : putriroroandini@gmail.com

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

☒ Skripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

ANALISIS DETEKSI EMOSI PADA TEKS BAHASA INDONESIA DAN BAHASA JAWA
DENGAN METODE CROSS-LINGUAL MENGGUNAKAN MODERN MULTILINGUAL
BERT

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.

Surabaya, 12 Januari 2026

Penulis

(Putri Roro Andini)

ABSTRAK

ANALISIS DETEKSI EMOSI PADA TEKS BAHASA INDONESIA DAN BAHASA JAWA DENGAN METODE CROSS-LINGUAL MENGUNAKAN MODERN MULTILINGUAL BERT

Oleh:

Putri Roro Andini

Deteksi emosi pada teks merupakan tantangan utama dalam *Natural Language Processing* (NLP), terutama untuk bahasa dengan sumber daya rendah seperti Bahasa Indonesia, Bahasa Sunda, dan Bahasa Jawa. Penelitian ini bertujuan mengevaluasi kinerja model berbasis Transformer dalam menyelesaikan tugas klasifikasi emosi *multi-label* pada skenario lintas bahasa. Model yang diuji meliputi ModernBERT dan mmBERT yang dilatih menggunakan dataset Bahasa Inggris dan Bahasa Sunda. Analisis data menunjukkan tantangan berupa ketimpangan kelas, sehingga penelitian ini menerapkan konfigurasi pembobotan (*class weighting*) pada fungsi *loss* BCEWithLogitsLoss untuk meningkatkan sensitivitas model terhadap kelas minoritas. Evaluasi performa dilakukan menggunakan metrik *macro F1-score*. Hasil eksperimen menunjukkan bahwa model ModernBERT mencapai performa optimal pada skenario *monolingual* Bahasa Inggris dengan *macro F1-score* sebesar 0.685. Sementara itu, model mmBERT menunjukkan keunggulan konsisten pada seluruh tugas lintas bahasa. Perbandingan pengaruh bahasa pelatihan mengungkapkan bahwa volume data Bahasa Inggris memberikan keunggulan pada pelatihan dibandingkan kedekatan rumpun antar Bahasa.

Kata Kunci: Deteksi Emosi, Klasifikasi Multi-label, Lintas Bahasa, ModernBERT, mmBERT, Ketimpangan Kelas.

ABSTRACT

ANALYSIS OF EMOTION DETECTION IN INDONESIAN AND JAVANESE TEXT WITH CROSS-LINGUAL METHOD USING MODERN MULTILINGUAL BERT

By:

Putri Roro Andini

Emotion detection in text is a major challenge in Natural Language Processing (NLP), especially for low-resource languages such as Indonesian, Sundanese, and Javanese. This study aims to evaluate the performance of Transformer-based models in solving multi-label emotion classification tasks in cross-language scenarios. The tested models include ModernBERT and mmBERT trained on English and Sundanese datasets. Data analysis revealed challenges in the form of class imbalance, so this study applied a class weighting configuration to the BCEWithLogitsLoss loss function to increase the model's sensitivity to minority classes. Performance evaluation was conducted using the macro F1-score metric. Experimental results show that the ModernBERT model achieves optimal performance in monolingual English scenarios with a macro F1-score of 0.685. Meanwhile, the mmBERT model shows consistent superiority across all cross-language tasks. Comparison of the effects of training languages reveals that the volume of English data provides an advantage in training compared to the proximity of the language family.

Kata kunci: Emotion Detection, Multi-label Classification, Cross-language, ModernBERT, mmBERT, Class Imbalance.

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	iii
PENGESAHAN TIM PENGUJI SKRIPSI.....	iv
PERNYATAAN KEASLIAN.....	v
ABSTRAK	vii
ABSTRACT.....	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xv
BAB I	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
BAB II.....	8
2.1 Tinjauan Penelitian Terdahulu.....	8
2.2 Dasar Teori	12
2.2.1 Emosi.....	12
2.2.2 Bahasa	13
2.2.3 Natural Language Processing (NLP).....	13
2.2.4 Cross Lingual Language.....	14
2.2.5 Transformers	14
2.2.6 BERT.....	17

2.2.7	ModernBERT	19
2.2.8	Modern Multilingual BERT	21
2.2.9	Dataset BRIGHTER	21
2.2.10	BCEWithLogitsLoss	22
2.2.11	Metrik Evaluasi	23
2.2.12	t-Distributed Stochastic Neighbor Embedding (t-SNE)	24
2.3	Integrasi Keilmuan	25
BAB III		27
3.1	Tahap Penelitian	27
3.2	Identifikasi Masalah	28
3.3	Studi Literatur	29
3.4	Pengumpulan Data	29
3.4.1	Bahasa Inggris	30
3.4.2	Bahasa Indonesia	36
3.4.3	Bahasa Jawa	40
3.4.4	Bahasa Sunda	45
3.5	Pre-Processing Data	50
3.5.1	Normalisasi	51
3.5.2	Tokenisasi	52
3.5.3	Padding dan Truncation	55
3.6	Pemodelan	55
3.7	Evaluasi dan Analisis Data	56
3.8	Skenario Eksperimen	57
3.9	Alat Penelitian	58
BAB IV		59
4.1	Pengumpulan Data	59

4.2	Preprocessing Data	65
4.2.1	Normalisasi.....	67
4.2.2	Pre-Tokenization (Splitting/Pemisahan)	71
4.2.3	Segmentasi Subword	83
4.2.4	Penambahan Token Khusus	89
4.2.5	Konversi Ke Token ID	98
4.3	Pemodelan.....	99
4.4	Evaluasi.....	102
4.5	Analisis Hasil.....	107
4.5.1	Perbandingan Performa ModernBERT dan mmBERT dalam Zero-Shot Cross-Lingual Transfer	107
4.5.2	Pengaruh Bahasa Pelatihan: Bahasa Inggris dan Bahasa Sunda	110
		111
BAB V		113
5.1	Kesimpulan	113
5.2	Saran	113
DAFTAR PUSTAKA		115

DAFTAR GAMBAR

Gambar 2. 1 Model Arsitektur Transformers (Vaswani et al., 2017)	15
Gambar 2.2 Arsitektur BERT	18
Gambar 2.3 Arsitektur Modern BERT	20
Gambar 3.1 Tahap Alur Penelitian.....	27
Gambar 3.2 Distribusi Label Data <i>Training</i> Bahasa Inggris.....	31
Gambar 3.3 Distribusi Kardinalitas Label Data <i>Train</i> Bahasa Inggris	32
Gambar 3.4 Distribusi Label Dataset <i>Dev</i> Bahasa Inggris.....	33
Gambar 3.5 Distribusi Kardinalitas Label Data <i>Dev</i> Bahasa Inggris.....	34
Gambar 3.6 Distribusi Label Dataset <i>Testing</i> Bahasa Inggris	35
Gambar 3.7 Distribusi Kardinalitas Label Data <i>Testing</i> Bahasa Inggris	36
Gambar 3.8 Distribusi Label Dataset <i>Dev</i> Bahasa Indonesia	37
Gambar 3.9 Distribusi Kardinalitas Label Data <i>Dev</i> Bahasa Indonesia	38
Gambar 3.10 Distribusi Label Dataset <i>Testing</i> Bahasa Indonesia	39
Gambar 3.11 Distribusi Kardinalitas Label Data <i>Testing</i> Bahasa Indonesia	40
Gambar 3.12 Distribusi Label Dataset <i>Dev</i> Bahasa Jawa	41
Gambar 3.13 Distribusi Kardinalitas Label Data <i>Dev</i> Bahasa Jawa.....	42
Gambar 3.14 Distribusi Label Dataset <i>Testing</i> Bahasa Jawa.....	44
Gambar 3.15 Distribusi Kardinalitas Label Data <i>Testing</i> Bahasa Jawa.....	44
Gambar 3.16 Distribusi Label Dataset <i>Training</i> Bahasa Sunda	46
Gambar 3.17 Distribusi Kardinalitas Label Data <i>Training</i> Bahasa Sunda	46
Gambar 3.18 Distribusi Label Dataset <i>Dev</i> Bahasa Sunda.....	48
Gambar 3.19 Distribusi Kardinalitas Label Data <i>Dev</i> Bahasa Sunda.....	49
Gambar 4.1 Alur <i>Preprocessing</i>	51
Gambar 4.2 Visualisasi Pairwise t-SNE Antar Bahasa.....	60
Gambar 4.3 Distribusi Panjang Token Data <i>Training</i> (Bahasa Inggris).....	61
Gambar 4.4 <i>Scatter Plot</i> Persebarang Panjang Token Bahasa Inggris	62
Gambar 4.5 Matriks Korelasi Antar Label Data <i>Train</i> Bahasa Inggris	63
Gambar 4.6 Distribusi Panjang Token Data <i>Training</i> (Bahasa Sunda)	64
Gambar 4.7 <i>Scatter Plot</i> Persebarang Panjang Token Bahasa Sunda.....	64
Gambar 4.8 Matriks Korelasi Antar Label Data <i>Train</i> Bahasa Sunda.....	65
Gambar 4.9 Hasil Pengujian Model Modern BERT	104

Gambar 4.10 Hasil Pengujian Model mmBERT.....	106
Gambar 4.11 Jarak Antar Bahasa.....	111



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	8
Tabel 3.1 Skema Dataset.....	29
Tabel 3.2 Sampel Dataset <i>Training</i> (Bahasa Inggris)	30
Tabel 3. 3 Sampel Dataset Dev (Bahasa Inggris)	32
Tabel 3.4 Sampel Dataset <i>Testing</i> (Bahasa Inggris)	34
Tabel 3.5 Sampel Dataset <i>Dev</i> (Bahasa Indonesia)	36
Tabel 3.6 Sampel Dataset <i>Testing</i> (Bahasa Indonesia)	38
Tabel 3.7 Sampel Dataset <i>Dev</i> (Bahasa Jawa)	40
Tabel 3.8 Sampel Dataset <i>Testing</i> (Bahasa Jawa).....	43
Tabel 3.9 Sampel Dataset <i>Train</i> (Bahasa Sunda)	45
Tabel 3.10 Sampel Dataset <i>Dev</i> (Bahasa Sunda)	47
Tabel 3.11 Perbandingan basis AutoTokenizer	50
Tabel 3.12 Perbandingan Spesifikasi Model.....	56
Tabel 3.13 Skenario Pengujian	57
Tabel 3.14 Spesifikasi Google Colab.....	58
Tabel 4.1 Teks Sampel Uji Tokenizer.....	66
Tabel 4.2 Hasil Normalisasi Data Sampel Uji	67
Tabel 4.3 Hasil <i>Pre-Tokenization</i> Data Sampel Uji.....	71
Tabel 4.4 Hasil <i>Sub-word Tokenization</i> Data Sampel Uji	83
Tabel 4.5 Hasil Penanganan Token Khusus Data Sampel Uji	91
Tabel 4.6 Pembagian Skenario Berdasar Bahasa dan Model.....	99
Tabel 4.7 Sampel Hasil Prediksi	101
Tabel 4.8 Hasil Pengujian Model ModernBERT	103
Tabel 4.9 Hasil Pengujian Model mmBERT	105
Tabel 4.10 Ringkasan Performa Terbaik	107
Tabel 4.11 Perbandingan Performa Berdasarkan Bahasa	112

DAFTAR PUSTAKA

- Aji, A. F. (2022). One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 7226–7249. <https://doi.org/10.18653/v1/2022.acl-long.500>
- Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt. Boston : Houghton Mifflin Harcourt.
- 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* (No. arXiv:2104.08200). arXiv. <https://doi.org/10.48550/arXiv.2104.08200>
- Chomsky, N. (1965). *Aspects of the Theory of Syntax* (50th ed.). The MIT Press. <http://www.jstor.org/stable/j.ctt17kk81z>
- Christen, P., Hand, D. J., & Kirielle, N. (2024). A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives. *ACM Computing Surveys*, 56(3), 1–24. <https://doi.org/10.1145/3606367>
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–8451. <https://doi.org/10.18653/v1/2020.acl-main.747>

- Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). *GoEmotions: A Dataset of Fine-Grained Emotions* (No. arXiv:2005.00547). arXiv. <https://doi.org/10.48550/arXiv.2005.00547>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200. <https://doi.org/10.1080/02699939208411068>
- Elkan, C. (2001). The Foundations of Cost-Sensitive Learning. *Proceedings of the Seventeenth International Conference on Artificial Intelligence: 4-10 August 2001; Seattle, 1*.
- Goodfellow, I., Courville, A., & Bengio, Y. (2016). *Deep learning*. The MIT Press.
- Herscovich, D., Frank, S., Lent, H., De Lhoneux, M., Abdou, M., Brandl, S., Bugliarello, E., Cabello Piqueras, L., Chalkidis, I., Cui, R., Fierro, C., Margatina, K., Rust, P., & Søgaard, A. (2022). Challenges and Strategies in Cross-Cultural NLP. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 6997–7013. <https://doi.org/10.18653/v1/2022.acl-long.482>
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261–266. <https://doi.org/10.1126/science.aaa8685>
- Hsu, T., Liu, C., & Lee, H. (2019). *Zero-shot Reading Comprehension by Cross-lingual Transfer Learning with Multi-lingual Language Representation*

- Model (No. arXiv:1909.09587). arXiv.
<https://doi.org/10.48550/arXiv.1909.09587>
- Huang, Z., & Cui, X. (2025). *PromotionGo at SemEval-2025 Task 11: A Feature-Centric Framework for Cross-Lingual Multi-Emotion Detection in Short Texts* (No. arXiv:2507.08499). arXiv.
<https://doi.org/10.48550/arXiv.2507.08499>
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). *Scaling Laws for Neural Language Models* (No. arXiv:2001.08361). arXiv.
<https://doi.org/10.48550/arXiv.2001.08361>
- Keung, P., Lu, Y., & Bhardwaj, V. (2020). *Adversarial Learning with Contextual Embeddings for Zero-resource Cross-lingual Classification and NER* (No. arXiv:1909.00153). arXiv. <https://doi.org/10.48550/arXiv.1909.00153>
- Kridalaksana, H. (2013). *Kamus Linguistik (Edisi Keempat)*. Gramedia Pustaka Utama. <https://books.google.co.id/books?id=gKNLDwAAQBAJ>
- Lauscher, A., Ravishankar, V., Vulić, I., & Glavaš, G. (2020). *From Zero to Hero: On the Limitations of Zero-Shot Cross-Lingual Transfer with Multilingual Transformers* (No. arXiv:2005.00633). arXiv.
<https://doi.org/10.48550/arXiv.2005.00633>
- Lin, Y.-H., Chen, C.-Y., Lee, J., Li, Z., Zhang, Y., Xia, M., Rijhwani, S., He, J., Zhang, Z., Ma, X., Anastasopoulos, A., Littell, P., & Neubig, G. (2019). Choosing Transfer Languages for Cross-Lingual Learning. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3125–3135. <https://doi.org/10.18653/v1/P19-1301>

- Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), 2579–2605.
- Marone, M., Weller, O., Fleshman, W., Yang, E., Lawrie, D., & Durme, B. V. (2025). *mmBERT: A Modern Multilingual Encoder with Annealed Language Learning* (No. arXiv:2509.06888). arXiv. <https://doi.org/10.48550/arXiv.2509.06888>
- Mohammad, S., & Kiritchenko, S. (2018). Understanding Emotions: A Dataset of Tweets to Study Interactions between Affect Categories. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA). <https://aclanthology.org/L18-1030/>
- Muhammad, S. H., Ousidhoum, N., Abdulmumin, I., Wahle, J. P., Ruas, T., Beloucif, M., Kock, C. de, Surange, N., Teodorescu, D., Ahmad, I. S., Adelani, D. I., Aji, A. F., Ali, F. D. M. A., Alimova, I., Araujo, V., Babakov, N., Baes, N., Bucur, A.-M., Bukula, A., ... Mohammad, S. M. (2025). *BRIGHTER: BRIdging the Gap in Human-Annotated Textual Emotion Recognition Datasets for 28 Languages* (No. arXiv:2502.11926). arXiv. <https://doi.org/10.48550/arXiv.2502.11926>
- Öhman, E., Pàmies, M., Kajava, K., & Tiedemann, J. (2020). XED: A Multilingual Dataset for Sentiment Analysis and Emotion Detection. *Proceedings of the 28th International Conference on Computational Linguistics*, 6542–6552. <https://doi.org/10.18653/v1/2020.coling-main.575>

- Ponti, E. M., Glavaš, G., Majewska, O., Liu, Q., Vulić, I., & Korhonen, A. (2020). XCOPA: A Multilingual Dataset for Causal Commonsense Reasoning. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2362–2376. <https://doi.org/10.18653/v1/2020.emnlp-main.185>
- Ruder, S., Vulić, I., & Søgaard, A. (2019). A Survey of Cross-lingual Word Embedding Models. *Journal of Artificial Intelligence Research*, 65, 569–631. <https://doi.org/10.1613/jair.1.11640>
- Rust, P., Pfeiffer, J., Vulić, I., Ruder, S., & Gurevych, I. (2021). *How Good is Your Tokenizer? On the Monolingual Performance of Multilingual Language Models* (No. arXiv:2012.15613). arXiv. <https://doi.org/10.48550/arXiv.2012.15613>
- Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, 23(7), 1307–1351. <https://doi.org/10.1080/02699930902928969>
- Seyeditabari, A., Tabari, N., & Zadrozny, W. (2018). *Emotion Detection in Text: A Review* (No. arXiv:1806.00674). arXiv. <https://doi.org/10.48550/arXiv.1806.00674>
- Shazeer, N. (2020). *GLU Variants Improve Transformer* (No. arXiv:2002.05202). arXiv. <https://doi.org/10.48550/arXiv.2002.05202>
- Sumekto, D. R., Ghozali, I., Yuwono, S. E., Santoso, G. B., & Tukiyo, T. (2022). Javanese Politeness Experience as Depicted in Its Speech Levels of the Transactional Communication. *Jurnal Humaniora*, 34(1), 36. <https://doi.org/10.22146/jh.65058>

- Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, 57, 117–126. <https://doi.org/10.1016/j.eswa.2016.03.028>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2019). *GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding* (No. arXiv:1804.07461). arXiv. <https://doi.org/10.48550/arXiv.1804.07461>
- Warner, B., Chaffin, A., Clavié, B., Weller, O., Hallström, O., Taghadouini, S., Gallagher, A., Biswas, R., Ladhak, F., Aarsen, T., Cooper, N., Adams, G., Howard, J., & Poli, I. (2024). *Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference* (No. arXiv:2412.13663). arXiv. <https://doi.org/10.48550/arXiv.2412.13663>
- Weiss, G. M. (2004). Mining with rarity: A unifying framework. *ACM SIGKDD Explorations Newsletter*, 6(1), 7–19. <https://doi.org/10.1145/1007730.1007734>
- Winata, G. I., Aji, A. F., Cahyawijaya, S., Mahendra, R., Koto, F., Romadhony, A., Kurniawan, K., Moeljadi, D., Prasajo, R. E., Fung, P., Baldwin, T., Lau, J. H., Sennrich, R., & Ruder, S. (2023). NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages. *Proceedings of the 17th Conference of the European Chapter of the Association for*

<https://doi.org/10.18653/v1/2023.eacl-main.57>

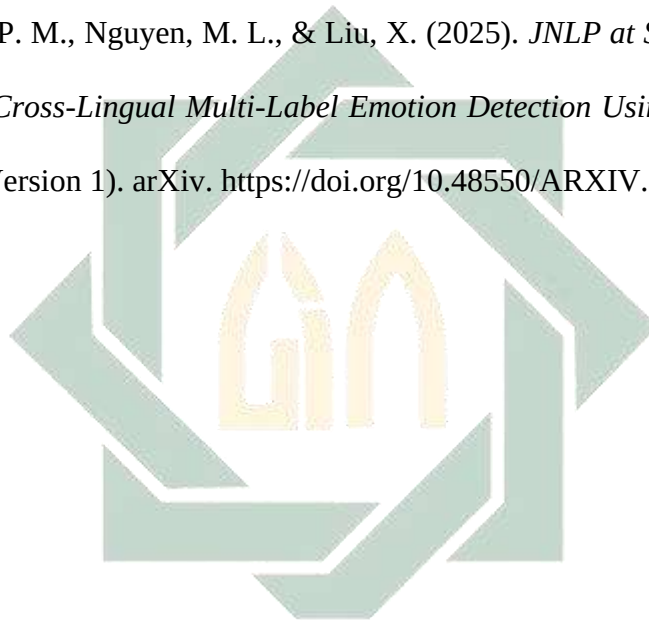
Winata, G. I., Madotto, A., Lin, Z., Liu, R., Yosinski, J., & Fung, P. (2021).

Language Models are Few-shot Multilingual Learners. *Proceedings of the 1st Workshop on Multilingual Representation Learning*, 1–15.

<https://doi.org/10.18653/v1/2021.mrl-1.1>

Xue, J., Nguyen, P. M., Nguyen, M. L., & Liu, X. (2025). *JNLP at SemEval-2025*

Task 11: Cross-Lingual Multi-Label Emotion Detection Using Generative Models (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2505.13244>



UIN SUNAN AMPEL
S U R A B A Y A