FINE-TUNING INDOBERT UNTUK KLASIFIKASI KATEGORI BERITA BERBAHASA INDONESIA

Diajukan Untuk Menyusun Skripsi
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI

Oleh:

Kiagus Muhammad Efan Fitriyan

NIM : 09021282126039

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2024

LEMBAR PENGESAHAN SKRIPSI

FINE-TUNING INDOBERT UNTUK KLASIFIKASI KATEGORI BERITA BERBAHASA INDONESIA

Oleh:

Kiagus Muhammad Efan Fitriyan

NIM: 09021282126039

Inderalaya, 27 Desember 2024

Pembimbing I,

Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Pembimbing II,

Mastura Diana Marieska, S.T., M.T.
NIP. 198603212018032001

Mengetahui,
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari jumat tangga tanggal 27 Desember 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama	:	Kiagus Muhammad Efan Fitriyan
Nim	:	09021282126039
Judul	:	*Fine-tuning IndoBERT* untuk Klasifikasi Kategori Berita Berbahasa Indonesia

dan dinyatakan **LULUS.**

1. Ketua Penguji

   Desty Rodiah, S.Kom., M.T.
   NIP. 198912212020122011 .....

2. Penguji I

   Muhammad Qurhanul Rizqie, S.Kom., M.T., Ph.D. .....
   NIP. 198712032022031006

3. Pembimbing I

   Novi Yusliani, S.Kom., M.T. .....
   NIP. 198211082012122001

4. Pembimbing II

   Mastura Diana Marieska, S.T., M.T. .....
   NIP. 198603212018032001

Mengetahui
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Kiagus Muhammad Efan Fitriyan

NIM : 09021282126039

Program Studi : Teknik Informatika

Judul Skripsi : *Fine-tuning IndoBERT* untuk Klasifikasi Kategori Berita Berbahasa Indonesia

Hasil pengecekan Software Turnitin : 6%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak mana pun.

Palembang, 30 Desember 2024

Penulis,

Kiagus Muhammad Efan Fitriyan

NIM. 09021282126039

MOTTO DAN PERSEMBAHAN

Motto:

"*Pengetahuan tanpa tindakan adalah sesuatu yang sia-sia, dan tindakan tanpa pengetahuan adalah kebodohan*"

- Ibnu Sina

Kupersembahkan Karya Tulis ini kepada:

- Allah SWT
- Orang Tua
- Keluarga Besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

The availability of Indonesian news articles on the internet has greatly increased, making it more challenging to recognize and categorize news accurately. Therefore, a solution to this issue is to develop a classification system for Indonesian news article categories. This research aims to classify Indonesian news category using fine-tuning on the pre-trained IndoBERT model. The dataset consists of 31,993 articles divided into five news categories: education, health, technology, sports, and automotive. Articles were collected from two of the largest and most trusted online news portals, kompas.com and detik.com, using web scraping method. The fine-tuning process was divided into 8 scenarios, which are combinations of dataset type configurations, learning rate, and batch size. Based on the test results, the highest accuracy was obtained in scenario 2, where the model trained with a learning rate of 2e-5 and batch size of 32, reaching an accuracy of 98.37%.

Keywords: *News Classification, Pre-trained Model, IndoBERT, Fine-tuning, Web Scraping, Accuracy*

ABSTRAK

Ketersediaan artikel berita berbahasa Indonesia yang tersebar di internet saat ini sudah sangat banyak yang mengakibatkan proses mengenali dan mengategorikan berita menjadi semakin sulit. Oleh karena itu, solusi untuk permasalahan mengategorikan artikel berita adalah mengembangkan sistem klasifikasi kategori berita berbahasa Indonesia. Penelitian ini bertujuan untuk mengklasifikasikan kategori berita berbahasa Indonesia dengan melakukan *fine-tuning pre-trained* model IndoBERT. *Dataset* yang digunakan berjumlah 31.993 terbagi menjadi 5 kategori berita yakni, pendidikan, kesehatan, teknologi, olahraga, dan otomotif. Artikel berita dikumpulkan dari dua portal berita *online* terbesar dan terpercaya yakni, kompas.com dan detik.com, menggunakan metode *web scraping*. Proses *fine-tuning* dibagi menjadi 8 skenario yang merupakan kombinasi dari konfigurasi jenis *dataset*, *learning rate* dan *batch size*. Berdasarkan hasil pengujian, akurasi tertinggi diperoleh pada skenario 2 yaitu, model yang dilatih dengan konfigurasi *learning rate* 2e-5 dan *batch size* 32 dengan akurasi sebesar 98,37%.

Kata Kunci: Klasifikasi Berita, *Pre-trained* Model, IndoBERT, *fine-tuning*, *Web Scraping*, Akurasi

KATA PENGANTAR

Puji dan syukur kepada Allah SWT yang telah memberikan nikmat sehat, iman dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini dibuat sebagai salah satu syarat untuk menyelesaikan Pendidikan program Strata-1 di Fakultas Ilmu Komputer Universitas Sriwijaya. Dalam proses pembuatan skripsi ini penulis menerima bimbingan, bantuan, semangat, maupun petunjuk dari banyak pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya sehingga, penulis dapat menyelesaikan skripsi ini dengan baik
2. Kedua orang tua dan keluarga yang telah mendoakan, memberi semangat dan motivasi untuk menyelesaikan skripsi ini.
3. Bapak Hadipurnawan Satria, Ph.D. selaku ketua jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Dr. M. Fachrurrozi, S.Si., M.T. selaku Dosen dan sekaligus pembimbing akademik yang telah memberikan banyak bantuan dan arahan kepada penulis selama perkuliahan.
5. Ibu Novi Yusliani, S.Kom., M.T. selaku Dosen Pembimbing skripsi I dan Ibu Mastura Diana Marieska, S.T., M.T. selaku Dosen Pembimbing skripsi II yang telah memberi bimbingan, arahan serta semangat kepada penulis dalam menyelesaikan skripsi ini.

6. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Saudari Yolendri Anisyahfitri yang telah memberikan bantuan, motivasi dan semangat kepada penulis.
8. Seluruh staf Administrasi dan Pegawai Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi.
9. Sahabat-sahabat penulis yang telah memberikan saran, kritik, dan motivasi kepada penulis.
10. Pihak-pihak lain yang tidak dapat penulis sebutkan satu-persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak sekali kekurangan karena kurangnya pengalaman dan pengetahuan penulis. Oleh karena itu penulis mengharapkan saran dan kritik yang membangun guna kemajuan penelitian selanjutnya. Semoga tugas akhir ini dapat bermanfaat. Terima Kasih.

Palembang, 30 Desember 2024
Penulis,

Kiagus Muhammad Efan Fitriyan

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1 Pendahuluan

Bab ini akan membahas mengenai landasan penelitian seperti, latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, dan sistematika penulisan. Pembahasan-pembahasan tersebut berfungsi sebagai kerangka dasar proses penelitian.

1.2 Latar Belakang Masalah

Ketersediaan informasi berita melalui internet telah menjadi faktor penting bagi banyak pengguna internet, meningkatkan kemungkinan mereka untuk mengakses berita dengan cepat dan mudah (Juarto, 2023). Berita memiliki informasi yang bermanfaat untuk berbagai tujuan, seperti analisis, pengambilan keputusan, atau sekadar mendapatkan pemahaman yang lebih mendalam mengenai suatu topik. Namun, seiring dengan meningkatnya jumlah berita yang tersedia, proses mengenali dan mengategorikan informasi menjadi semakin sulit. Salah satu sistem yang bisa digunakan untuk mengategorikan informasi adalah sistem klasifikasi teks. Klasifikasi teks merupakan cara yang bisa digunakan untuk mengenali kategori berita atau informasi yang sangat banyak secara efektif dan membantu dalam mengurangi kompleksitas data yang sangat besar (Gunawan & Santoso, 2021). Klasifikasi teks mampu menganalisis pola dan fitur dalam teks secara otomatis, seperti kata kunci, struktur kalimat, dan konteks suatu berita, untuk mengenali kategori yang sesuai dengan teks yang diklasifikasikan, sehingga proses

klasifikasi teks ini berguna untuk mengenali dan memilah sebuah berita yang sesuai dengan kebutuhan pengguna.

Salah satu pendekatan yang bisa digunakan untuk klasifikasi teks berita adalah dengan melakukan *fine-tuning* pada *pre-trained* model BERT (Sun et al., 2019). Penelitian ini menggunakan *pre-trained* model, BERT, untuk melakukan klasifikasi teks berita. BERT adalah model bahasa yang sangat baik untuk memahami konteks dan hubungan antar kata-kata dalam teks (Jin et al., 2020). IndoBERT merupakan *pre-trained* model BERT yang telah dilatih dengan miliaran data kosakata dari bahasa Indonesia. Model IndoBERT terbukti mengungguli beberapa model dasar lainnya dalam berbagai tugas *natural language processing* dalam konteks bahasa Indonesia, termasuk model fastText dan *multilingual* BERT (Faisal & Mahendra, 2022).

Meskipun model *multilingual* BERT telah dilatih dengan berbagai bahasa selain bahasa Inggris, termasuk bahasa Indonesia, namun dalam beberapa penelitian, IndoBERT masih menunjukkan keunggulannya. Berdasarkan penelitian yang dilakukan oleh (Nugroho et al., 2021), dalam kasus analisis sentimen pada ulasan aplikasi *mobile* berbahasa Indonesia, model IndoBERT mengungguli model *multilingual* BERT, pada *batch size* 32 IndoBERT mampu mendapatkan akurasi sebesar 82% sedangkan *multilingual* BERT sebesar 78% (Nugroho et al., 2021).

Penelitian yang dilakukan oleh (Juarto, 2023) membandingkan kinerja model IndoBERT dengan model *machine learning* seperti XGBoost, *random forest*, dan *light gradient boosting*. Penelitian ini menggunakan *dataset AGNews* dan hasil

pengujian menunjukkan bahwa IndoBERT mencapai tingkat akurasi tertinggi sebesar 95%, sementara *light gradient boosting*, hanya mencapai akurasi 92%. Selain itu, evaluasi terhadap beberapa model *natural language processing* lainnya seperti XLMNet, XLM Roberta, dan *multilingual* BERT menunjukkan bahwa IndoBERT mencapai akurasi lebih tinggi sebesar 94,5%, dengan *training loss* dan *validation loss* yang lebih rendah, serta membutuhkan waktu komputasi yang lebih singkat. Keunggulan IndoBERT dalam klasifikasi berita dikarenakan IndoBERT memiliki pemahaman yang lebih baik terhadap struktur dan konteks bahasa Indonesia (Hutama & Suhartono, 2022). Penelitian lain yang dilakukan oleh (Faisal & Mahendra, 2022) menunjukkan bahwa IndoBERT memiliki akurasi 87,02% dan *F1-score* 60,12% melampaui model-model lainnya dalam kasus deteksi kesalahan informasi COVID-19 pada *tweet* berbahasa Indonesia. Hasil pengujian membuktikan bahwa model IndoBERT memiliki kinerja terbaik dalam tugas klasifikasi teks. Namun, untuk meningkatkan akurasi dan generalisasi model, disarankan untuk memperluas *dataset* dengan data yang lebih beragam. Penambahan data dari berbagai sumber, topik, dan gaya penulisan akan membantu model menangani variasi bahasa yang lebih luas, sehingga mampu memberikan prediksi yang lebih akurat dan dapat diandalkan dalam berbagai konteks.

Penelitian sebelumnya mengungkapkan bahwa meskipun IndoBERT memiliki keunggulan dalam berbagai tugas klasifikasi teks, terdapat keterbatasan pada generalisasi model akibat kurangnya keragaman *dataset*, seperti yang dijelaskan oleh (Hutama & Suhartono, 2022) dan (Faisal & Mahendra, 2022). Keduanya merekomendasikan perluasan *dataset* untuk meningkatkan akurasi dan

kemampuan generalisasi model. Fokus penelitian ini adalah pengoptimalan IndoBERT untuk kategorisasi berita berbahasa Indonesia dengan *dataset* yang mencakup lima kategori yaitu, pendidikan, kesehatan, teknologi, olahraga, dan otomotif. Pada penelitian ini tidak hanya mengevaluasi kinerja model pada satu jenis *dataset* tertentu, tetapi juga menguji kemampuan IndoBERT untuk digeneralisasi ke berbagai topik berita dengan *dataset* yang lebih beragam.

Berdasarkan ulasan yang telah dibahas, penelitian ini bertujuan untuk membangun sistem klasifikasi teks berita berbahasa Indonesia dengan melakukan *fine-tuning* pada *pre-trained* model IndoBERT dan menguji kemampuan IndoBERT untuk digeneralisasi ke berbagai topik berita dengan *dataset* yang lebih beragam.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka didapat rumusan masalah pada penelitian ini yaitu:

1. Bagaimana mengembangkan sistem klasifikasi kategori berita berbahasa Indonesia dengan melakukan *fine-tuning* model IndoBERT?
2. Bagaimana kinerja sistem klasifikasi kategori berita berbahasa Indonesia dengan melakukan *fine-tuning* model IndoBERT berdasarkan tingkat akurasi?

1.4 Tujuan Penulisan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sistem yang dapat mengklasifikasikan kategori berita berbahasa Indonesia dengan *fine-tuning* model IndoBERT.

2. Mengetahui kinerja sistem klasifikasi kategori berita berbahasa Indonesia dengan *fine-tuning* model IndoBERT berdasarkan tingkat akurasi.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Perangkat lunak yang dihasilkan dapat digunakan untuk melakukan klasifikasi kategori berita berbahasa Indonesia.
2. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan teknik klasifikasi teks dalam bahasa Indonesia.

1.6 Batasan Penelitian

Batasan masalah penelitian ini adalah sebagai berikut:

1. Berita yang digunakan merupakan berita berbahasa Indonesia.
2. Kategori teks berita yang digunakan, yaitu pendidikan, kesehatan, teknologi, olahraga, dan otomotif.
3. *Pre-trained* model yang digunakan adalah IndoBERT-*Base Phase* 1.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini mengikuti standar operasional penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yakni:

BAB I. PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan yang digunakan dalam penyusunan laporan akhir ini.

BAB II. KAJIAN LITERATUR

Bab ini menjelaskan mengenai landasan teori yang digunakan dalam penelitian. Dalam bab ini membahas mengenai kajian literatur, seperti penjelasan mengenai *pre-trained* model IndoBERT, *fine-tuning*, serta penelitian sebelumnya yang berkaitan dengan penelitian ini

BAB III. METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan proses yang dilakukan selama penelitian seperti, metode pengumpulan data, pelatihan model, hingga metode dalam perancangan perangkat lunak. Setiap tahapan penelitian dijelaskan secara rinci sesuai dengan kerangka kerja yang ditetapkan.

1.8 Kesimpulan

Pada bab pendahuluan yang telah dijelaskan sebelumnya, dapat disimpulkan bahwa penelitian ini membahas *fine-tuning pre-trained* model IndoBERT untuk klasifikasi teks berita berbahasa Indonesia.

DAFTAR PUSTAKA

Adhim, F. I., Martin, R. F., Budiprayitno, S., & Rahayu, L. P. (2022). Development of Employee Payroll System using Rational Unified Process (RUP) on Odoo Platform. *Applied Technology and Computing Science Journal*, *5*(1), 36–43. https://doi.org/10.33086/atcsj.v5i1.3696

Akhmad, E. P. A. (2023). Analisis Sentimen Ulasan Aplikasi DLU Ferry Pada Google Play Store Menggunakan Bidirectional Encoder Representations from Transformers. *Jurnal Aplikasi Pelayaran Dan Kepelabuhanan*, *13*(2), 104–112. https://doi.org/10.30649/japk.v13i2.94

Arase, Y., & Tsujii, J. (2019). Transfer Fine-Tuning: A BERT Case Study. *EMNLP 2019*. https://doi.org/10.18653/v1/D19-1542

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

Faisal, D. R., & Mahendra, R. (2022). Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets. *Elsevier Journal*. http://arxiv.org/abs/2206.15359

Fajri Akbar, A., Santoso, H. B., Oktavia, P., Putra, H., & Yudhoatmojo, S. B. (2021). User Perception Analysis of Online Learning Platform "Zenius" During the Coronavirus Pandemic Using Text Mining Techniques. In *Journal of Information System)* (Vol. 17, Issue 2).

Gunawan, K. I., & Santoso, J. (2021). Multilabel Text Classification Menggunakan SVM dan Doc2Vec Classification Pada Dokumen Berita Bahasa Indonesia. *Journal of Information System,Graphics, Hospitality and Technology*, *3*(01), 29–38. https://doi.org/10.37823/insight.v3i01.126

Gururangan, S., Marasovic, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). *Don't Stop Pretraining: Adapt Language Models to Domains and Tasks*. https://github.com/allenai/

Hafiz, Y. A., & Sudarmilah, E. (2023). *Implementasi Web Scraping pada Portal Berita Online*.

Hossain, A., Karimuzzaman, M., Hossain, M. M., & Rahman, A. (2021). Text mining and sentiment analysis of newspaper headlines. *Information (Switzerland)*, *12*(10). https://doi.org/10.3390/info12100414

Hutama, L. B., & Suhartono, D. (2022). Indonesian Hoax News Classification with Multilingual Transformer Model and BERTopic. *Informatica (Slovenia)*, *46*(8), 81–90. https://doi.org/10.31449/inf.v46i8.4336

Jin, D., Jin, Z., Zhou, J. T., & Szolovits, P. (2020). Is BERT Really Robust? A Strong Baseline for Natural Language Attack on Text Classification and Entailment. *AAAI Press*, *34*(05). https://doi.org/https://doi.org/10.1609/aaai.v34i05.6311

Juarto, B. (2023). Indonesian News Classification Using IndoBert. *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, *2023*(2), 454. www.ijisae.org

Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(4), 887–894. https://doi.org/10.25126/jtiik.1148315

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*. http://arxiv.org/abs/2011.00677

Kruchten, P. (1999). *Le Rational Unified Process ®*.

Miyajiwala, A., Ladkat, A., Jagadale, S., & Joshi, R. (2022). *On Sensitivity of Deep Learning Based Text Classification Algorithms to Practical Input Perturbations*. https://doi.org/10.1007/978-3-031-10464-0_42

Nanda, R., Haerani, E., Gusti, S. K., & Ramadhani, S. (2022). Klasifikasi Berita Menggunakan Metode Support Vector Machine. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, *5*(2).

Nguyen, D. Q., Vu, T., & Nguyen, A. T. (2020). *BERTweet: A pre-trained language model for English Tweets*. http://arxiv.org/abs/2005.10200

Nugroho, K. S., Sukmadewa, A. Y., DW, H. W., Bachtiar, F. A., & Yudistira, N. (2021). BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews. *Association for Computing Machinery*. https://research.google/teams/brain.

Perwitasari, R., Afwani, R., & Anjarwani, S. E. (2020). *Penerapan Metode Rational Unified Process (RUP) dalam Pengembenagan Sistem Informasi Medical Check Up pada Citra Medical Centre*. http://jtika.if.unram.ac.id/index.php/JTIKA/

Priyambodo, A., & Prihati, P. (2020). *Evaluasi Ekstraksi Fitur Klasifikasi Teks untuk Peningkatan Akurasi Klasifikasi menggunakan Naive Bayes*. https://doi.org/https://doi.org/10.51903/elkom.v13i1.277

Putra, T. I. Z. M., Suprapto, S., & Bukhori, A. F. (2022). Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia. *Jurnal Ilmu Siber Dan Teknologi Digital*, *1*(1), 1–28. https://doi.org/10.35912/jisted.v1i1.1509

Rahmawati, A., Alamsyah, A., & Romadhony, A. (2022). Hoax News Detection Analysis using IndoBERT Deep Learning Methodology. *2022 10th International Conference on Information and Communication Technology (ICoICT)*, 368–373. https://doi.org/10.1109/ICoICT55009.2022.9914902

Ramli, N. E., Yahya, Z. R., & Said, N. A. (2022). Confusion Matrix as Performance Measure for Corner Detectors. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *29*(1), 256–265. https://doi.org/10.37934/araset.29.1.256265

Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(4), 377–383. https://doi.org/10.47233/jteksis.v5i4.913

Sahria, Y. (2020). *Implementasi Teknik Web Scraping pada Jurnal SINTA*. http://sinta2.ristekdikti.go.id/journals/detail

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). *How to Fine-Tune BERT for Text Classification?* http://arxiv.org/abs/1905.05583

Valero-Carreras, D., Alcaraz, J., & Landete, M. (2023). Comparing two SVM models through different metrics based on the confusion matrix. *Computers & Operations Research*, *152*, 106131. https://doi.org/10.1016/j.cor.2022.106131

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding*. http://arxiv.org/abs/2009.05387

Yefferson, D. Y., Lawijaya, V., & Girsang, A. S. (2024). Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news. *IAES International Journal of Artificial Intelligence*, *13*(2), 1911–1922. https://doi.org/10.11591/ijai.v13.i2.pp1913-1924