STANCE CLASSIFICATION DALAM ISU PEMBANGUNAN IKN MENGGUNAKAN SVM DAN INDOSBERT

SKRIPSI

UIN SUNAN AMPEL
S U R A B A Y A

Disusun Oleh :

ALDI PUTRA WIDYANSYAH
09010621002

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	: Aldi Putra Widyansyah
NIM	: 09010621002
Program Studi	: Sistem Informasi
Angkatan	: 2021

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "STANCE CLASSIFICATION DALAM ISU PEMBANGUNAN IKN MENGGUNAKAN SVM DAN INDOSBERT". 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, 16 Juni 2025

Yang Menyatakan,

Aldi Putra Widyansyah

NIM. 09010621002

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh,

Nama	:	Aldi Putra Widyansyah
NIM	:	09010621002
Judul	:	STANCE DETECTION DALAM ISU PEMBANGUNAN IKN MENGGUNAKAN SVM DAN INDOSBERT

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 Aldi Putra Widyansyah ini telah dipertahankan di depan tim penguji skripsi di Surabaya, 16 Juni 2025

Mengesahkan,

Dewan Penguji

Dosen Penguji 1	Dosen Penguji 2
Bayu Adhi Nugroho, Ph.D	Khalid, M.Kom
NIP. 197905182014031001	NIP. 197906092014031002

Dosen Penguji 3	Dosen Penguji 4
Dr. Eng. Anang Kunaefi, M.Kom	Subhan Nooriansyah, M.Kom
NIP. 197911132014031001	NIP. 199012282020121010

Mengetahui,

Dekan Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya

Dr. A. Saepul 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 : Aldi Putra Widyansyah
NIM : 09010621002
Fakultas/Jurusan : Sains dan Teknologi / Sistem Informasi
E-mail address : aldi.widyansyah@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 :

STANCE CLASSIFICATION DALAM ISU PEMBANGUNAN IKN MENGGUNAKAN SVM DAN INDOSBERT

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif 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, 11 Juli 2025
Penulis

Aldi Putra Widyansyah

ABSTRAK

STANCE CLASSIFICATION DALAM ISU PEMBANGUNAN IKN MENGGUNAKAN SVM DAN INDOSBERT

Oleh:

Aldi Putra Widyansyah

Isu pembangunan Ibu Kota Nusantara (IKN) menjadi perhatian utama dalam kebijakan publik pemerintahan di Indonesia. Perpindahan ibu kota dan pembangunan ibu kota baru diharapkan dapat mendukung pemerataan pembangunan dan meningkatkan kualitas hidup masyarakat. Namun, di balik harapan tersebut, terdapat beragam pandangan dan keberpihakan pendapat yang muncul dalam masyarakat. Penelitian ini bertujuan untuk melakukan stance detection dan klasifikasi terhadap isu pembangunan IKN dengan memanfaatkan algoritma Support Vector Machine (SVM) dan model bahasa IndoSBERT. Model IndoSBERT digunakan untuk menghasilkan embedding teks dari tweet terkait IKN, yang kemudian diolah menggunakan SVM untuk stance classification. Stance classification dibagi menjadi tiga kelas, yaitu Mendukung, Netral, dan Menolak. Proses augmentasi data dilakukan untuk menyeimbangkan distribusi kelas, menghasilkan 464 tweet per kelas, sebelum ekstraksi fitur menggunakan IndoSBERT. Hasil evaluasi menunjukkan bahwa model SVM dengan konfigurasi validasi silang 7-fold (K=7) mencapai performa terbaik, dengan akurasi per fold mencapai 72% - 79% dan akurasi rata-rata sebesar 75%, F1-score sebesar 75%. Hasil ini menegaskan bahwa pendekatan berbasis IndoSBERT dan SVM efektif untuk mengklasifikasikan sikap masyarakat terhadap IKN.

Kata Kunci: ibu kota nusantara, IndoSBERT, SVM, *stance detection*, pembangunan ibu kota

ABSTRACT

STANCE CLASSIFICATION IN THE ISSUE OF IKN DEVELOPMENT USING SVM AND INDOSBERT

By:

Aldi Putra Widyansyah

Development of the Nusantara Capital or Ibu Kota Nusantara (IKN) has become a primary focus in Indonesia's public policy. The development and construction of the new capital are expected to promote equitable development and increasing the quality of economy and life for the population. However, behind these expectations, diverse perspectives and biases in public opinion have emerged. This study aims to perform stance detection and classification on the issue of IKN development by utilizing the Support Vector Machine (SVM) algorithm and the IndoSBERT language model. IndoSBERT was employed to generate text embeddings from tweets related to IKN, which were then processed using SVM for stance classification. The stance classification was categorized into three classes: Support, Neutral, and Oppose. Data augmentation was conducted to balance the class distribution, resulting in 464 tweets per class, prior to feature extraction using IndoSBERT. The evaluation results demonstrate that the SVM model with a 7-fold cross-validation configuration (K=7) achieved the best performance, with per-fold accuracy ranging from 72% to 79%, an average accuracy of 75%, and an F1-score of 75%. These findings confirm that the IndoSBERT and SVM-based approach is effective in classifying public stances toward IKN development.

Keywords: Nusantara Capital, IndoSBERT, SVM, stance detection, capital development

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR PUSTAKA

Abdelrahim, Elsaid Md., Hasan Hashim, El-Sayed Atlam, Radwa Ahmed Osman, and Ibrahim Gad. "TMS: Ensemble Deep Learning Model for Accurate Classification of Monkeypox Lesions Based on Transformer Models with SVM." *Diagnostics* 14, no. 23 (November 23, 2024): 2638. https://doi.org/10.3390/diagnostics14232638.

Akbar, Yuma, Anggit Nur Hannaa Regita, Sugiyono, and Tri Wahyudi. "Analisa Sentimen Pada Media Sosial'X' Pencarian Keyword ChatGPT Menggunakan Algoritma K-Nearest Neighbors (KNN)." *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi* 5, no. 3 (September 30, 2024): 3291–3305. https://doi.org/10.35870/jimik.v5i3.1016.

Alturayeif, Nora, Hamzah Luqman, and Moataz Ahmed. "A Systematic Review of Machine Learning Techniques for Stance Detection and Its Applications." *Neural Computing and Applications* 35, no. 7 (March 2023): 5113–44. https://doi.org/10.1007/s00521-023-08285-7.

Amato, Alberto, and Vincenzo Di Lecce. "Data Preprocessing Impact on Machine Learning Algorithm Performance." *Open Computer Science* 13, no. 1 (July 17, 2023): 20220278. https://doi.org/10.1515/comp-2022-0278.

Ariestya, Winda Widya, Ida Astuti, and I Made Wiryana. "Preprocessing For Crawler Of Short Message Social Media." In *2018 Third International Conference on Informatics and Computing (ICIC)*, 1–6. Palembang, Indonesia: IEEE, 2018. https://doi.org/10.1109/IAC.2018.8780451.

Augenstein, Isabelle, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. "Stance Detection with Bidirectional Conditional Encoding." In *Proceedings of the 2016 Conference on Empirical Methods in*

Natural Language Processing, 876–85. Austin, Texas: Association for Computational Linguistics, 2016. https://doi.org/10.18653/v1/D16-1084.

Bengio, Y., & Grandvalet, Y. (2004). No Unbiased Estimator of the Variance of K-Fold Cross-Validation. *The Journal of Machine Learning Research*, *5*(1089–1105).

Brin, Sergey, and Lawrence Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks and ISDN Systems* 30, no. 1–7 (April 1998): 107–17. https://doi.org/10.1016/S0169-7552(98)00110-X.

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, *2*, 121–167. https://doi.org/10.1023/A:1009715923555

Burnham, Michael. "Stance Detection: A Practical Guide to Classifying Political Beliefs in Text." *Political Science Research and Methods*, September 19, 2024, 1–18. https://doi.org/10.1017/psrm.2024.35.

Chicco, Davide, and Giuseppe Jurman. "The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation." *BMC Genomics* 21, no. 1 (December 2020): 6. https://doi.org/10.1186/s12864-019-6413-7.

Darwish, Kareem, Peter Stefanov, Michaël Aupetit, and Preslav Nakov. "Unsupervised User Stance Detection on Twitter." *Proceedings of the International AAAI Conference on Web and Social Media* 14 (May 26, 2020): 141–52. https://doi.org/10.1609/icwsm.v14i1.7286.

Denny, Matthew James, and Arthur Spirling. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To

Do About It." *SSRN Electronic Journal*, 2017. https://doi.org/10.2139/ssrn.2849145.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North*, 4171–86. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. https://doi.org/10.18653/v1/N19-1423.

Dey, Kuntal, Ritvik Shrivastava, and Saroj Kaushik. "Topical Stance Detection for Twitter: A Two-Phase LSTM Model Using Attention." arXiv, January 9, 2018. https://doi.org/10.48550/arXiv.1801.03032.

Ghosh, Shalmoli, Prajwal Singhania, Siddharth Singh, Koustav Rudra, and Saptarshi Ghosh. "Stance Detection in Web and Social Media: A Comparative Study," 11696:75–87, 2019. https://doi.org/10.1007/978-3-030-28577-7_4.

Glandt, Kyle, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. "Stance Detection in COVID-19 Tweets." In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1596–1611. Online: Association for Computational Linguistics, 2021. https://doi.org/10.18653/v1/2021.acl-long.127.

Hamad, Omama, Ali Hamdi, Sayed Hamdi, and Khaled Shaban. "StEduCov: An Explored and Benchmarked Dataset on Stance Detection in Tweets towards Online Education during COVID-19 Pandemic." *Big Data and Cognitive Computing* 6, no. 3 (August 22, 2022): 88. https://doi.org/10.3390/bdcc6030088.

Hochreiter, Sepp, and Jürgen Schmidhuber. “Long Short-Term Memory.” *Neural Computation* 9, no. 8 (November 1, 1997): 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

Hatta, H. R., Mokoginta, D., Munawar, Z., Chafid, N., Nooriansyah, S., Suparman, A., Fitriyadi, F., Tanti, L., Saputro, A. K., & Hendrawan, L. (2024). *Kecerdasan buatan*. Yayasan Cendikia Mulia Mandiri.

Ipmawati, Joang, Saifulloh Saifulloh, and Kusnawi Kusnawi. “Analisis Sentimen Tempat Wisata Berdasarkan Ulasan Pada Google Maps Menggunakan Algoritma Support Vector Machine: Sentiment Analysis of Tourist Attractions Based on Reviews on Google Maps Using the Support Vector Machine Algorithm.” *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 4, no. 1 (January 15, 2024): 247–56. https://doi.org/10.57152/malcom.v4i1.1066.

Joachims, Thorsten. “Making Large Scale SVM Learning Practical.” *Forschungsberichte Des Lehrstuhls VIII*. Universität Dortmund, October 29, 1999. https://doi.org/10.17877/DE290R-5098.

Küçük, Dilek, and Fazli Can. “A Tutorial on Stance Detection.” In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 1626–28. Virtual Event AZ USA: ACM, 2022. https://doi.org/10.1145/3488560.3501391.

———. “Stance Detection: A Survey.” *ACM Computing Surveys* 53, no. 1 (January 31, 2021): 1–37. https://doi.org/10.1145/3369026.

Kuo, Kuan-Hung, Ming-Hung Wang, Hung-Yu Kao, and Yu-Chen Dai. “Advancing Stance Detection of Political Fan Pages: A Multimodal Approach.” In *Companion Proceedings of the ACM Web Conference 2024*,

702–5. Singapore Singapore: ACM, 2024. https://doi.org/10.1145/3589335.3651467.

Lee, Sarah, Zidian Xie, Emily Xu, Yihan Shao, Deborah J. Ossip, and Dongmei Li. "Public Perceptions of the FDA's Marketing Authorization of Vuse on Twitter/X." *Frontiers in Public Health* 11 (November 3, 2023): 1280658. https://doi.org/10.3389/fpubh.2023.1280658.

Li, Yingjie, and Cornelia Caragea. "Multi-Task Stance Detection with Sentiment and Stance Lexicons." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 6298–6304. Hong Kong, China: Association for Computational Linguistics, 2019. https://doi.org/10.18653/v1/D19-1657.

Li, Yingjie, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea. "P-Stance: A Large Dataset for Stance Detection in Political Domain." In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2355–65. Online: Association for Computational Linguistics, 2021. https://doi.org/10.18653/v1/2021.findings-acl.208.

Li, Yingjie, Chenye Zhao, and Cornelia Caragea. "Improving Stance Detection with Multi-Dataset Learning and Knowledge Distillation." In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 6332–45. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021. https://doi.org/10.18653/v1/2021.emnlp-main.511.

Lindelöf, Gabriel, Talayeh Aledavood, and Barbara Keller. "Dynamics of the Negative Discourse Toward COVID-19 Vaccines: Topic Modeling Study

and an Annotated Data Set of Twitter Posts." *Journal of Medical Internet Research* 25 (April 12, 2023): e41319. https://doi.org/10.2196/41319.

Lorge, Isabelle, Li Zhang, Xiaowen Dong, and Janet B. Pierrehumbert. "STEntConv: Predicting Disagreement with Stance Detection and a Signed Graph Convolutional Network." arXiv, 2024. https://doi.org/10.48550/ARXIV.2403.15885.

Mohammad, Saif M., Parinaz Sobhani, and Svetlana Kiritchenko. "Stance and Sentiment in Tweets." arXiv, 2016. https://doi.org/10.48550/ARXIV.1605.01655.

Murtfeldt, Ryan, Naomi Alterman, Ihsan Kahveci, and Jevin D. West. "RIP Twitter API: A Eulogy to Its Vast Research Contributions." arXiv, April 10, 2024. https://doi.org/10.48550/arXiv.2404.07340.

Nainggolan, E. U. P. (2022). Urgensi pemindahan ibu kota negara. *Direktorat Jenderal Kekayaan Negara*. Diakses pada 25 September 2024, https://www.djkn.kemenkeu.go.id/kanwil-kalbar/baca-artikel/14671/Urgensi-Pemindahan-Ibu-Kota-Negara.html

Putra, Yanfa Adi, and Masayu Leylia Khodra. "Deep Learning and Distributional Semantic Model for Indonesian Tweet Categorization." In *2016 International Conference on Data and Software Engineering (ICoDSE)*, 1–6. Denpasar, Indonesia: IEEE, 2016. https://doi.org/10.1109/ICODSE.2016.7936108.

Rahadika Diana, Kadek Denaya, and Masayu Leylia Khodra. "IndoSBERT: Enhancing Indonesian Sentence Embeddings with Siamese Networks Fine-Tuning." In *2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 1–6. Lombok, Indonesia: IEEE, 2023. https://doi.org/10.1109/ICAICTA59291.2023.10390469.

Rahmatulloh, Alam, and Rohmat Gunawan. "Web Scraping with HTML DOM Method for Data Collection of Scientific Articles from Google Scholar." *Indonesian Journal of Information Systems* 2, no. 2 (February 26, 2020): 95–104. https://doi.org/10.24002/ijis.v2i2.3029.

Reimers, Nils, and Iryna Gurevych. "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3980–90. Hong Kong, China: Association for Computational Linguistics, 2019. https://doi.org/10.18653/v1/D19-1410.

Revo. (2024). Bukan Jawa! Ini 10 provinsi dengan pertumbuhan ekonomi tertinggi di RI. *CNBC Indonesia*. Diakses pada 25 September 2024, https://www.cnbcindonesia.com/research/20240206100323-128-512191/bukan-jawa-ini-10-provinsi-dengan-pertumbuhan-ekonomi-tertinggi-di-ri

Rodriguez, J.D., A. Perez, and J.A. Lozano. "Sensitivity Analysis of K-Fold Cross Validation in Prediction Error Estimation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, no. 3 (March 2010): 569–75. https://doi.org/10.1109/TPAMI.2009.187.

Sennrich, R., Haddow, B., & Birch, A. (2015). *Improving Neural Machine Translation Models with Monolingual Data* (Version 4). arXiv. https://doi.org/10.48550/ARXIV.1511.06709

Septiani, R. K., Anggraeni, S., & Saraswati, S. D. (2022). Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes. *Teknika*, *16*(02). https://doi.org/10.5281/zenodo.7535887

Setiawan, Esther Irawati, and Ika Lestari. "Stance Classification Pada Berita Berbahasa Indonesia Berbasis Bidirectional LSTM." *Journal of Intelligent System and Computation* 3, no. 1 (April 1, 2021): 41–48. https://doi.org/10.52985/insyst.v3i1.148.

Shaw, Peter, Jakob Uszkoreit, and Ashish Vaswani. "Self-Attention with Relative Position Representations." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 464–68. New Orleans, Louisiana: Association for Computational Linguistics, 2018. https://doi.org/10.18653/v1/N18-2074.

Siddiqua, Umme Aymun, Abu Nowshed Chy, and Masaki Aono. "Tweet Stance Detection Using an Attention Based Neural Ensemble Model." In *Proceedings of the 2019 Conference of the North*, 1868–73. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. https://doi.org/10.18653/v1/N19-1185.

Syifa, Maharani Aulia, and Dewi Retno Sari Saputro. "Stance Detection Dengan Algoritme Gated Recurrent Unit (GRU)." *Prosiding Seminar Nasional Matematika Dan Statistika; Vol 3 No 01 (2023): Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya*, August 26, 2023. https://jurnal.fmipa.unmul.ac.id/index.php/SNMSA/article/view/1204.

Taranukhin, Maksym, Vered Shwartz, and Evangelos Milios. "Stance Reasoner: Zero-Shot Stance Detection on Social Media with Explicit Reasoning." arXiv, 2024. https://doi.org/10.48550/ARXIV.2403.14895.

Upadhyaya, Apoorva, Marco Fisichella, and Wolfgang Nejdl. "A Multi-Task Model for Sentiment Aided Stance Detection of Climate Change Tweets."

Proceedings of the International AAAI Conference on Web and Social Media 17 (June 2, 2023): 854–65. https://doi.org/10.1609/icwsm.v17i1.22194.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need." arXiv, 2017. https://doi.org/10.48550/ARXIV.1706.03762.

Wang, Bin, and C.-C. Jay Kuo. "SBERT-WK: A Sentence Embedding Method by Dissecting BERT-Based Word Models." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020): 2146–57. https://doi.org/10.1109/TASLP.2020.3008390.

Zaky, Damar, and Ade Romadhony. "An LSTM-Based Spell Checker for Indonesian Text." In *2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, 1–6. Yogyakarta, Indonesia: IEEE, 2019. https://doi.org/10.1109/ICAICTA.2019.8904218.

Zhang, X., Zhao, J., & LeCun, Y. (2015). *Character-level Convolutional Networks for Text Classification* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.1509.01626

Zhu, Yiming, Peixian Zhang, Ehsan-Ul Haq, Pan Hui, and Gareth Tyson. "Can ChatGPT Reproduce Human-Generated Labels? A Study of Social Computing Tasks." arXiv, April 22, 2023. https://doi.org/10.48550/arXiv.2304.10145.

Zhuo, T. Y., Yang, Z., Sun, Z., Wang, Y., Li, L., Du, X., Xing, Z., & Lo, D. (2023). *Source Code Data Augmentation for Deep Learning: A Survey* (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2305.19915