TOPIC MODELING SKRIPSI PROGRAM STUDI SISTEM INFORMASI MENGGUNAKAN KOMBINASI BERTOPIC DAN INDOSBERT

SKRIPSI

Disusun oleh :

NUR HUDA RIYANTONI
09020620037

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
SURABAYA
2024

LEMBAR PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama	:	NUR HUDA RIYANTONI
NIM	:	09020620037
Program Studi	:	Sistem Informasi
Angkatan	:	2020

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "TOPIC MODELING SKRIPSI PROGRAM STUDI SISTEM INFORMASI MENGGUNAKAN KOMBINASI BERTOPIC DAN INDOSBERT". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat maka bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 10 Mei 2024

Yang menyatakan,

Nur Huda Riyantoni

NIM. 09020620037

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA	:	NUR HUDA RIYANTONI
NIM	:	09020620037
JUDUL	:	*TOPIC MODELING* SKRIPSI PROGRAM STUDI SISTEM INFORMASI MENGGUNAKAN KOMBINASI BERTOPIC DAN INDOSBERT

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 10 Mei 2024

Dosen Pembimbing 1	Dosen Pembimbing 2
Khalid, M. Kom	Subhan Nooriansyah, M.Kom.
NIP. 197906092014031002	NIP. 199012282020121010

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Nur Huda Riyantoni ini telah dipertahankan

Di depan tim penguji skripsi

Di Surabaya, 20 Mei 2024

Mengesahkan,

Dewan Penguji

Dosen Penguji 1

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

Dosen Penguji 2

Ahmad Yusuf, M. Kom
NIP. 199001202014031003

Dosen Penguji 3

Khalid, M. Kom
NIP. 197906092014031002

Dosen Penguji 4

Subhan Nooriansyah, M.Kom
NIP. 199012282020121010

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya

Dr. H. 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 : NUR HUDA RIYANTONI
NIM : 09020620037
Fakultas/Jurusan : SAINS DAN TEKNOLOGI/ SISTEM INFORMASI
E-mail address : riyantoni2772@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 :

TOPIC MODELING SKRIPSI PROGRAM STUDI SISTEM INFORMASI MENGGUNAKAN KOMBINASI BERTOPIC DAN INDOSBERT

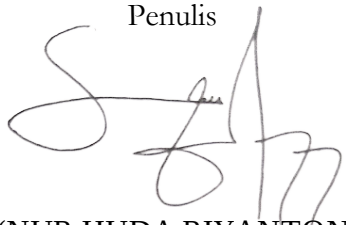
beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media 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, 10 Juni 2024

Penulis



(NUR HUDA RIYANTONI)

ABSTRAK

TOPIC MODELING SKRIPSI PROGRAM STUDI SISTEM INFORMASI MENGGUNAKAN KOMBINASI BERTOPIC DAN INDOSBERT

Oleh:

Nur Huda Riyantoni

Program Studi Sistem Informasi merupakan salah satu program studi tingkat sarjana yang populer di kalangan mahasiswa di Indonesia. Dalam menyelesaikan Program Studi tingkat sarjana salah satu syaratnya adalah menyelesaikan skripsi. Bersamaan dengan berjalannya waktu, jumlah dokumen skripsi yang telah dihasilkan mahasiswa program studi sistem informasi semakin meningkat. Banyaknya skripsi yang dihasilkan tentunya akan memakan waktu pembaca untuk mengidentifikasi topik yang ada dalam skripsi. Oleh karena itu, perlu dilakukan penerapan *topic modeling* untuk mengidentifikasi berbagai topik yang ada dalam kumpulan skripsi tersebut. Penelitian ini bertujuan untuk mengetahui topik penelitian skripsi mahasiswa program studi sistem informasi. Pada penelitian ini, *topic modeling* dilakukan menggunakan kombinasi BERTopic dan IndoSBERT. IndoSBERT digunakan dalam tahap *embeddings* BERTopic untuk menggantikan SBERT. Penelitian dilakukan dengan beberapa skenario nilai parameter UMAP, HDBSCAN dan CountVectorizer yang kemudian dievaluasi menggunakan *topic coherence* dan *topic diversity*. Hasil penelitian menunjukkan bahwa, skenario yang menghasilkan 7 topik memiliki nilai *topic coherence* dan *topic diversity* terbaik daripada skenario lainnya. 7 topik tersebut antara lain adalah evaluasi kualitas *website*, kepuasan pelanggan, tata kelola TI, segmentasi pelanggan, rancang bangun sistem informasi, sistem pendukung keputusan, dan manajemen risiko TI.

Kata kunci: *Topic modeling*, BERtopic, IndoSBERT, Skripsi, Sistem Informasi

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

ABSTRACT

TOPIC MODELING THESIS INFORMATION SYSTEMS STUDY PROGRAM USING A COMBINATION OF BERTOPIC AND INDOSBERT

By:

Nur Huda Riyantoni

The Information Systems Study Program is one of the undergraduate study programs that is popular among students in Indonesia. In completing an undergraduate study program, one of the requirements is to complete a thesis. As time goes by, the number of thesis documents produced by students in the information systems study program is increasing. The large number of theses produced will of course take the reader's time to identify the topics in the thesis. Therefore, it is necessary to apply topic modelling to identify the various topics in the thesis collection. This research aims to determine the thesis research topic of students in the information systems study program. In this research, topic modeling was carried out using a combination of BERTopic and IndoSBERT. IndoSBERT is used in the BERTopic embeddings stage to replace SBERT. The research was carried out with several scenarios of UMAP, HDBSCAN and CountVectorizer parameter values which were then evaluated using topic coherence and topic diversity. The research results showed that the scenario that produced 7 topics had the best topic coherence and topic diversity scores compared to other scenarios. These 7 topics include website quality evaluation, customer satisfaction, IT governance, customer segmentation, information system design, decision support systems, and IT risk management.

Kata kunci: *Topic modeling*, BERtopic, IndoSBERT, Thesis, Information System

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

DAFTAR ISI

UIN SUNAN AMPEL
S U R A B A Y A

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

UIN SUNAN AMPEL
S U R A B A Y A
http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

UIN SUNAN AMPEL
S U R A B A Y A

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

DAFTAR GAMBAR

UIN SUNAN AMPEL
S U R A B A Y A
http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

UIN SUNAN AMPEL
S U R A B A Y A

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

DAFTAR TABEL

UIN SUNAN AMPEL
S U R A B A Y A
http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

UIN SUNAN AMPEL
S U R A B A Y A

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

DAFTAR PUSTAKA

A. Yani, D. D., Pratiwi, H. S., & Muhardi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, *7*(4), 257. https://doi.org/10.26418/justin.v7i4.30930

Abuzayed, A., & Al-Khalifa, H. (2021). BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique. *Procedia Computer Science*, *189*, 191–194. https://doi.org/10.1016/j.procs.2021.05.096

Alamsah, A. S. (2018). *Implementasi Sistem Temu Kembali Informasi Untuk Pencarian Buku Pada Toko Buku Online Menggunakan Metode Vector Space Model*. Universitas Muhammadiyah Gresik.

Albab, M. U., P, Y. K., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, *20*(2), Article 2. https://doi.org/10.26623/transformatika.v20i2.5374

Alfanzar, A. I. (2019). *Topic Modelling Skripsi Menggunakan Metode Latent Dirichlet Allocation*. UIN Sunan Ampel Surabaya.

Allaoui, M., Kherfi, M. L., & Cheriet, A. (2020). *Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study* (hlm. 317–325).

Amy. (2023, Maret 28). Hierarchical Topic Model for Airbnb Reviews. *Medium*.

Anggraini, E. (2020). *Latent Dirichlet Allocation Untuk Pemodelan Topik Abstrak Dokumen Skripsi (Studi Kasus: Abstrak Dokumen Skripsi Mahasiswa Statistika Uii Tahun Angkatan 2011-2015)*. UNIVERSITAS ISLAM INDONESIA YOGYAKARTA.

Axelborn, H., & Berggren, J. (2023). *Topic Modeling for Customer Insights A Comparative Analysis of LDA and BERTopic in Categorizing Customer Calls*. UMEA University.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*.

Bourequat, W., & Mourad, H. (2021). Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine. *International Journal of Advances in Data and Information Systems*, *2*(1), 36–44. https://doi.org/10.25008/ijadis.v2i1.1216

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. https://doi.org/10.48550/ARXIV.1810.04805

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2019). *Topic Modeling in Embedding Spaces*. https://doi.org/10.48550/ARXIV.1907.04907

Dihni, V. A. (2021). *Apa Program Studi Paling Diminati Mahasiswa Indonesia? | Databoks*. Databoks.

Dimitriadis, N. S. (2020). *Applying Topic Modelling Algorithms on Twitter messages in Greek language*. https://ikee.lib.auth.gr/record/324006/files/Dimitriadis-2158.pdf

Efimov, V. (2023, September 16). *Large Language Models: SBERT — Sentence-BERT*. Medium. https://towardsdatascience.com/sbert-deb3d4aef8a4

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*, 886498. https://doi.org/10.3389/fsoc.2022.886498

Firdaus, A., & Firdaus, W. I. (2021). *Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi: (Sebuah Ulasan)*. *13*(1).

Giri. (2020, November 6). *Topic Model Evaluation*. HDS. https://highdemandskills.com/topic-model-evaluation/

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. https://doi.org/10.48550/ARXIV.2203.05794

Grootendorst, M. (2024). *Documents—BERTopic*.

Herwinsyah, H. (2023). Pemodelan Topik Dalam Al-Qur'an Menggunakan Library Bertopic Pada Model Bahasa Bert. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, *14*(2), 319–328. https://doi.org/10.24176/simet.v14i2.9900

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, *78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Joachims, T. (1996). *A Probabilistic AnalysisToexf tthCeaRteogcocrhizioatAiolngorithm with TFIDF for*.

Kholilulloh, M. (2014). *Studi Tentang Kecenderungan Kajian Skripsi Pai Iain Sunan Ampel Tahun 2007-2012a*. UIN Sunan Ampel Surabaya.

Kumar, D. V., & Chadha, A. (2012). *Mining Association Rules in Student's Assessment Data*. *9*(5).

Listari. (2019). *Topic Modeling Menggunakan Latent Dirchlect Allocation (Part 1): Pre-processing Data dengan Python*. Medium.

Mandar, G., & Gunawan, G. (2017). Peringkasan dokumen berita Bahasa Indonesia menggunakan metode Cross Latent Semantic Analysis. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, *3*(2), 94. https://doi.org/10.26594/register.v3i2.1161

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

Mansurova, M. (2023, September 9). *Topics per Class Using BERTopic*. Medium. https://towardsdatascience.com/topics-per-class-using-bertopic-252314f2640

Mastery, T. (2023). *Bertopic – A Must Read Comprehensive Guide*. DotCom magazine. https://dotcommagazine.com/2023/07/bertopic-a-must-read-comprehensive-guide/

McInnes, L. (2018). *Basic UMAP Parameters—Umap 0.5 documentation*. https://umap-learn.readthedocs.io/en/latest/parameters.html

McInnes, L., Healy, J., & Astels, S. (2016). *Parameter Selection for HDBSCAN**—Hdbscan 0.8.1 documentation*. https://hdbscan.readthedocs.io/en/latest/parameter_selection.html

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (arXiv:1802.03426). arXiv. http://arxiv.org/abs/1802.03426

Mifrah, S. (2020). Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using COVID'19 Corpus. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 5756–5761. https://doi.org/10.30534/ijatcse/2020/231942020

Octavianto, A. W. (2023, Juli 21). *Topic Modeling dan BERTopic: Menggali Lebih Dalam Data Teks untuk Penemuan yang Lebih Kaya*. Medium.

Ogunleye, B., Maswera, T., Hirsch, L., Gaudoin, J., & Brunsdon, T. (2023). Comparison of Topic Modelling Approaches in the Banking Context. *Applied Sciences*, *13*(2), Article 2. https://doi.org/10.3390/app13020797

Pardede, D. L. C., & Waskita, M. A. I. (2023). Analisis Pemodelan Topik Untuk Ulasan Tentang Peduli Lindungi. *Jurnal Ilmiah Informatika Komputer*, *28*(1), 17–26. https://doi.org/10.35760/ik.2023.v28i1.7925

Patmawati, P., & Yusuf, M. (2021). Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter oleh Pejabat Negara. *Building of Informatics, Technology and Science (BITS)*, *3*(3), 122–129. https://doi.org/10.47065/bits.v3i3.1012

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.

Pedro, J. (2022, Januari 10). *Understanding Topic Coherence Measures*. Medium. https://towardsdatascience.com/understanding-topic-coherence-measures-4aa41339634c

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

Pradana, R. O. (2023). *Analisis tren topik sistem informasi di Indonesia dari perspektif Topic Modeling menggunakan LDA (Latent Dirichlet Allocation)*. UIN Sunan Ampel Surabaya.

Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. ANDI Yogyakarta.

Priyatna, G. G. (2022). *Pemodelan Topik Terkait Ulasan Video Game Dengan Genre Battle Royale Menggunakan Metode Bertopic Dengan Fitur Guided Topic Modelling*. Universitas Islam Negeri Syarif Hidayatullah Jakarta.

Putra, A. E. (2018). *Pengaruh Seleksi Fitur Chi-Square Terhadap Kinerja Algoritma Naive Bayes Classifier Pada Analisis Sentimen Dokumen*. Universitas Islam Negeri Syarif Hidayatullah Jakarta.

Putra, K. B., & Kusumawardani, R. P. (2017). Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). *Jurnal Teknik ITS*, *6*(2), A446-450. https://doi.org/10.12962/j23373539.v6i2.23205

Qiao, R. (2019). *Yelp Review Rating Prediction: Sentiment Analysis and the Neighborhood-Based Recommender*. UNIVERSITY OF CALIFORNIA.

Rahadika Diana, K. D., & Khodra, M. L. (2023). IndoSBERT: Enhancing Indonesian Sentence Embeddings with Siamese Networks Fine-tuning. *2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 1–6. https://doi.org/10.1109/ICAICTA59291.2023.10390469

Rahmatulloh, A., & Gunawan, R. (2020). Web Scraping with HTML DOM Method for Data Collection of Scientific Articles from Google Scholar. *Indonesian Journal of Information Systems*, *2*(2), 95–104. https://doi.org/10.24002/ijis.v2i2.3029

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (arXiv:1908.10084). arXiv. https://doi.org/10.48550/arXiv.1908.10084

Rijcken, E. (2023). *C_v Topic Coherence Explained*. Medium.

Saket, S. (2020, Januari 12). Count Vectorizers vs TFIDF Vectorizers| Natural Language Processing. *Artificial Coder*.

Salsabila, N. P. (2022). *Implementasi Deep Neural Network dalam Perancangan Respons Chatbot dengan Menggunakan Pendekatan Natural Language Processing*.

Samsir, S., Saragih, R. S., Subagio, S., Aditiya, R., & Watrianthos, R. (2023). BERTopic Modeling of Natural Language Processing Abstracts: Thematic Structure and Trajectory. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *7*(3), Article 3.

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

Satriajati, S., Panuntun, S. B., & Pramana, S. (2020). Implementasi Web Scraping Dalam Pengumpulan Berita Kriminal Pada Masa Pandemi Covid-19. *Seminar Nasional Official Statistics*, *2020*(1), Article 1. https://doi.org/10.34123/semnasoffstat.v2020i1.578

Septiyan, D. (2020). *IMPLEMENTASI METODE NAIVE BAYES CLASSIFIER TERHADAP ANALISIS SENTIMEN KOMENTAR PADA MEDIA SOSIAL*. Institut Informatika & Bisnis Darmajaya.

Shalahuddin, R. (2020). *Ridhwan102/Python-Spell-Checker-Bahasa-Indonesia* [Python]. (Original work published 2020)

Sihombing, Eirene. (2014). *Penerapan Analisis Sentimen dengan Metode Naïve Bayes pada Klasifikasi Data Teks*. Universitas Padjadjara.

Stewart, G., & Al-Khassaweneh, M. (2022). An Implementation of the HDBSCAN* Clustering Algorithm. *Applied Sciences*, *12*(5), 2405. https://doi.org/10.3390/app12052405

Susilo, A. (2023). *Analisis Sentimen Sara Pada Tweet Berbahasa Indonesia Menggunakan Indobert Dan Support Vector Machine (Svm)*. UIN Sunan Ampel Surabaya.

Tan, P.-N. (2006). *Introduction to data mining*. Pearson Addison Wesley.

Thamrin, H., Oktafiani, D., Rasyid, I. I., & Fauzi, I. M. (2024). Classification of SWOT Statements Employing BERT Pre-Trained Model Embedding. *Jurnal Sistem Informasi Bisnis*, *14*(2), 143–152. https://doi.org/10.21456/vol14iss2pp143-152

Tran, T. (2023, November 6). Topic Modelling: Crafting an LDA Model with Python for Analyzing Dialogue in the ‘Friends’ Sitcom. *Medium*.

Udit. (2023, Januari 1). The Jensen-Shannon Divergence: A Measure of Distance Between Probability Distributions. *Medium*.

Utomo, M. N. Ya. (2017, November 24). Spell Checker untuk Deteksi dan Perbaikan Typo Bahasa Indonesia. *YasirUtomo*.

van der Maaten, L., Postma, E., & Herik, H. (2007). Dimensionality Reduction: A Comparative Review. *Journal of Machine Learning Research - JMLR*, *10*.

Vasudeva Raju, S., Kumar Bolla, B., Nayak, D. K., & Kh, J. (2022). Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings. *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, 1–6. https://doi.org/10.1109/I2CT54291.2022.9824873

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/

Vijayarani, D. S., & Ilamathi, J. (2015). *Preprocessing Techniques for Text Mining—An Overview*. *5*.

Wibisono, O., Septiandri, A. A., & Najogie, R. D. (2024). *Assessing the Impact of ESG-Related News on Stock Trading in the Indonesian Market: A Text Similarity Framework Approach*.

Wisnu Setyawan, A. (2021). *Implementasi Algoritma Latent Dirichlet Allocation untuk Topic Modeling Terhadap Data Twitter Terkait Pandemi Covid-19*. https://kc.umn.ac.id/id/eprint/17957/

Zvornicanin, E. (2021, Desember 7). *When Coherence Score Is Good or Bad in Topic Modeling? | Baeldung on Computer Science*. https://www.baeldung.com/cs/topic-modeling-coherence-score

UIN SUNAN AMPEL
S U R A B A Y A

http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/http://digilib.uinsby.ac.id/