**MATA KULIAH NEURAL LANGANGUAGE PROCESSING (NLP)**

**TUGAS UJIAN TENGAH SEMESTER (UTS)**



**Oleh:**

**Asep Ridwan Hidayat**

**231012050036**

**PROGRAM STUDI TEKNIK INFORMATIKA S-2**

**PROGRAM PASCASARJANA**

**UNIVERSITAS PAMULANG**

**TANGERANG SELATAN**

**2025**

# DAFTAR ISI

[LEMBAR PERSETUJUAN PROPOSAL TESIS i](#_Toc184631815)

[LEMBAR PERNYATAAN TESIS ii](#_Toc184631816)

[KATA PENGANTAR iii](#_Toc184631817)

[ABSTRAK iv](#_Toc184631818)

[ABSTRACT v](#_Toc184631819)

[DAFTAR ISI vi](#_Toc184631820)

[BAB I PENDAHULUAN 1](#_Toc184631821)

[1.1 Latar Belakang (Time New roman, 12, Bold) 1](#_Toc184631822)

[1.2 Permasalahan Penelitian 1](#_Toc184631823)

[1.2.1 Identifikasi Masalah 1](#_Toc184631824)

[1.2.2 Ruang Lingkup Masalah 1](#_Toc184631825)

[1.2.3 Rumusan Masalah 2](#_Toc184631826)

[1.3 Tujuan dan Manfaat Penelitian 2](#_Toc184631827)

[1.4 Sistematika Penulisan 2](#_Toc184631828)

BAB II [LANDASAN TEORI DAN KERANGKA PEMIKIRAN 4](#_Toc184631830)

[2.1 Tinjauan Pustaka 4](#_Toc184631831)

[2.2 Landasan Teori 4](#_Toc184631832)

[2.2.1 *Artificial Inteligence* 4](#_Toc184631833)

[BAB III METODOLOGI 9](#_Toc184631839)

[3.1 Analisis Kebutuhan 9](#_Toc184631840)

[3.2 Perancangan Penelitian 9](#_Toc184631841)

[3.3 Teknik Analisis 9](#_Toc184631842)

[3.3.1 Teknik Analisis 9](#_Toc184631843)

[3.3.2 Evaluasi Model 9](#_Toc184631844)

[DAFTAR PUSTAKA 11](#_Toc184631845)

**DAFTAR TABEL**

[Tabel 2.1 Implementasi IOT](#_Toc37836972) 11

[Tabel 2.2 Daftar Simbol *Flowchart*](#_Toc37836972) 28

[Tabel 2.3 *Use Case Diagram*](#_Toc37836973) 28

Daftar tabel sesuai dengan bab dan urutan tabel di bab tersebut

Contoh :

1. Tabel di bab 2 dan tabel kedua maka menjadi Tabel 2.2
2. Tabel di bab 3 dan tabel pertama maka menjadi tabel 3.1

Lalu di ikuti keterangan tabel terkait apa

**DAFTAR GAMBAR**

Gambar 2.1 Konsep IOT 19

Gambar 2.2 Jenis-jenis ESP 8266 22

Gambar 2.3 *Real Time Clock* 23

Gambar 2.4 Breadboard 24

Daftar gambar sesuai dengan bab dan urutan gambar di bab tersebut

Contoh :

1. Gambar di bab 2 dan tabel kedua maka menjadi Gambar 2.2
2. Gambar di bab 3 dan tabel pertama maka menjadi Gambar 3.1

Lalu di ikuti keterangan Gambar terkait apa

**(diuat Secara otomatis seperti daftar isi)**

# TUGAS PERTEMUAN 1

## 10 jurnal terkait perkembangan NLP terkini ( Lima Tahun Terakhir)

| **No** | **Penulis** | **Tahun** | **Judul Penelitian** | **Model** | ***Accuracy*** | **Sumber** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Muhammad Fathan Fauzan, Rahmi Imanda, Muhammad Adryan Hasbi | 2025 | Designing an Chatbot with NLP Technology in a Website-Based New Student Admission Information System | *Chatbot, NLP* | First, blackbox testing shows that the system functions well in responding to messages sent through the chatbot on the website, both from text that matches the intention and text that is abstract and does not match the pattern, with an accuracy rate of 87.5%. | Journal of Applied Informatics and Computing (JAIC) |
| 2 | Mukhlis Amien | 2023 | Sejarah dan Perkembangan Teknik *Natural Language Processing* (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia | *NLP* | Studi ini juga menggali aplikasi NLP dalam industri dan penelitian bahasa Indonesia serta mengidentifikasi tantangan dan peluang dalam penelitian dan pengembangan NLP bahasa Indonesia. | ELANG: Journal of Interdisciplinary Research  E-ISSN: 3025-2482 |
| 3 | Andre Farhan Saputra, Kecitaan Harefa | 2025 | Penerapan Metode Natural Language Processing (Nlp) Dalam Implementasi Asisten Virtual Chatbot Dengan Memanfaatkan Api Chatgpt Dan Gradio App | *Chatbot, NLP* | The research results showed that from a questionnaire of 24 respondents, ChatGPT users who directly tested Chillbot provided positive feedback, with a user satisfaction rate of 87.17%. | Building of Informatics, Technology and Science (BITS) |
| 4 | Muhammad Rofiq Sudrajat, Muhammad Zakariyah | 2024 | Penerapan Natural Language Processing dan Machine Learning untuk Prediksi Stres Siswa SMA Berdasarkan Analisis Teks | NLP, Regression, Naive Bayes, Random Forest, dan Support Vector Machine (SVM) | Model Machine Learning yang diujicobakan antara lain Logistic Regression, Naive Bayes, Random Forest, dan Support Vector Machine (SVM). Hasil dari eksperimen menunjukkan bahwa model Naive Bayes yang menggunakan  fitur Bigram mencapai akurasi tertinggi yaitu 95,6%, dengan model lainnya mencapai sekitar 93%. | MDPI,Applied Science |
| 5 | Muhammad Yusuf, Indah Purnama Sari, Virda Kristy | 2024 | Sistem Pakar Mencegah Stunting dengan Menentukan Gizi Anak Menggunakan Natural Language Processing (NLP) | *NLP* | The strength of the NLP algorithm lies in its ability to understand user queries based on context, resulting in relevant and responsive solutions. The testing results indicate a system accuracy rate of 0.9756 or 97%, achieved through valuations using a dataset of user queries under various test scenarios. | Jurnal JTIK (Jurnal Teknologi Informasi dan komuninknasi)  OI: <https://doi.org/10.35870/>  jtik.v9i3.3614 |
| 6 | Xinyu Fu | 2024 | Natural Language Processing in Urban Planning: A Research Agenda | *NLP* | The results reveal that existing research is primarily exploratory with a fragmented research landscape. Future studies should focus on sharing data, benchmarking NLP techniques, fostering collaborative research tailored to planning, and addressing ethical implications to harness NLP’s full potential in planning | [https://journals.sagepub.com/ doi/10.117](https://journals.sagepub.com/%20doi/10.117) |
| 7 | Ghofrane Merhbene, Alexandre Puttick, Mascha Kurpicz-Briki | 2024 | Investigating machine learning and natural language processing techniques applied for detecting eating disorders: a systematic literature review | *NLP* | Investigasi pada jurnal ini mencakup empat area utama: (a) analisis metadata dari makalah yang diterbitkan, (b) pemeriksaan ukuran dan topik spesifik dari kumpulan data yang digunakan, (c) tinjauan penerapan teknik pembelajaran mesin dalam mendeteksi gangguan makan dari teks, dan terakhir (d) evaluasi model yang digunakan, dengan fokus pada kinerja, keterbatasan, dan potensi risiko yang terkait dengan metodologi saat ini. | <https://www.frontiersin.org/journals/>  psychiatry/articles/10.3389  /fpsyt.2024.1319522/full |
| 8 | Panteleimon Krasadakis, Evangelos Sakkopoulos, Vassilios S. Verykios | 2024 | *A Survey on Challenges and Advances in Natural Language Processing with a Focus on Legal Informatics and Low-Resource Languages* | NLP | the jurnal conducted an extensive literature review of NLP research focused on legislative documents. We present the current state-of-the-art NLP tasks related to Law Consolidation, highlighting the challenges that arise in low-resource languages. Our goal is to outline the difficulties faced by this field and the methods that have been developed to overcome them | https://www.mdpi.com/2079-9292/13/3/648 |
| 9 | M Raihan | 2024 | *Dynamic Topic Modelling* Menggunakan *BERTOPIC* Dalam Pemilihan Presiden Tahun 2019 | Bertopic | Peneliti mencoba menganalisis topik apa saja yang dihasilkan dari *tweet* yang diunggah oleh masyarakat menjelang Pemilu 2019 dan disertai dengan evolusi topiknya dari waktu ke waktu. Metode pemodelan topik yang akan digunakan kali ini adalah *BERTopic*. Metode pemodelan topik ini di dasari *sentence embedding* dengan salah satu jenis arsitektur *neural network* yaitu *Siamese network* sehingga metode ini dapat mengelompokkan kata sesuai konteksnya dalam suatu kalimat. Metode *BERTopic* ini juga dilengkapi dengan fitur *Dynamic Topic Modelling* yaitu metode pemodelan topik yang dilanjutkan dengan mengevolusi setiap topiknya dari waktu ke waktu. Dengan data *tweet* yang ada, metode *BERTopic* mampu menghasilkan topik-topik yang ada dengan baik, hal ini dapat dibuktikan dengan hasil evaluasi dari nilai koheren yang dihasilkan yaitu 0.71. Topik yang dihasilkan juga relevan dan dapat dibuat narasinya. | <https://repository.uinjkt.ac.id/> dspace/handle/123456789/ 81888 |
| 10 | Fitria | 2025 | Penerapan Metode NLP pada Chatbot Output Suara | NLP Chatbot | Integrasi NLP dengan output suara pada chatbot, training data, evaluasi model | Jurnal portal publikasi |

## 10 teknologi NLP yang yang telah diadopsi oleh Industri/ Government / Perusahaan TI

| **No** | **Penulis/Sumber** | **Tahun** | **Judul Penelitian** | **Model** | **Pembahasan** | **Sumber** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Eric W.T. Ngai, Ariel K.H. Lui, Brian C.W. Kei | 2024 | Natural language processing in government applications: A literature review and case analysis | *NLP* | Studi tinjauan literatur dan analisis kasus penggunaan NLP dalam aplikasi pemerintahan, termasuk efisiensi dokumen dan layanan publik. | Emerald insight  Discover Journals, Books & Case Studies |
| 2 | Yunqing Jiang, Patrick Cheong-Iao Pang, Dennis Wong ,Ho Yin Kan | 2023 | Natural Language Processing Adoption in Governments and Future Research Directions: A Systematic Review | *NLP* | Tinjauan sistematis adopsi NLP di pemerintahan, membahas aplikasi, manfaat, dan tantangan implementasi NLP di sektor publik. | <https://www.mdpi.com>  /2076-3417/  13/22/12346 |
| 3 | Dr. Jagreet Kaur Gill | 2025 | Natural Language Processing in Government: Complete Guide | *Chatbot, NLP* | Penjelasan peran NLP dalam transformasi pemerintahan menuju e-governance, termasuk chatbot, analisis sentimen, dan penerjemahan Bahasa | XenonStack Blog  (<https://www.xenonstack.com/>  blog/nlp-in-government) |
| 4 | Cogent Infotech | 2024 | Role of NLP in the Public Sector | NLP | Aplikasi NLP untuk menjawab pertanyaan publik, peningkatan kepolisian, analisis media sosial, dan ekstraksi informasi penting. | https://www.cogentinfo.com/resources/nlp-in-the-public-sector |
| 5 | SAS Whitepaper | 2023 | Revolutionizing Government Communication with NLP Applications | *NLP* | Fokus pada chatbot, analisis sentimen, penerjemahan bahasa, ekstraksi informasi, dan pengenalan suara dalam komunikasi pemerintah. | https://www.sas.com/content/dam/SAS/documents/briefs/solution-brief/en/natural-language-processing-government-113025.pdf |
| 6 | INA Solutions Blog | 2023 | Revolutionizing Government Communication with NLP Applications | *NLP, Chatbot* | Fokus pada chatbot, analisis sentimen, penerjemahan bahasa, ekstraksi informasi, dan pengenalan suara dalam komunikasi pemerintah. | https://ina-solutions.com/resources/2023/12/01/natural-language-processing-government-applications/ |
| 7 | Deloitte Insights | 2019 | Natural Language Processing Examples in Government Data | *NLP* | Contoh konkret penggunaan NLP untuk analisis data tak terstruktur di pemerintahan AS, termasuk deteksi pola dan kepatuhan regulasi. | https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/natural-language-processing-examples-in-government-data.html |
| 8 | Emerald Insigh | 2024 | *Natural Language Processing in Government Applications* | NLP | Studi komprehensif mengenai penggunaan NLP di pemerintahan, termasuk pengelolaan data, analisis kebijakan, dan pelayanan publik. | https://www.emerald.com/insight/content/doi/10.1108/imds-07-2024-0711/full/html |
| 9 | IBM Research | 2021 | *NLP in IT Companies: Enhancing Customer Support and Automation* | NLP | Aplikasi NLP dalam perusahaan IT untuk chatbot, analisis sentimen pelanggan, dan otomatisasi proses bisnis. (berdasarkan pengetahuan umum) |  |
| 10 | Microsoft Research | 2022 | Leveraging NLP for Enterprise Knowledge Management | NLP | Penggunaan NLP untuk ekstraksi informasi dan manajemen pengetahuan di perusahaan teknologi besar. (berdasarkan pengetahuan umum) |  |

# PERTEMUAN 3

## 2.1. Cari 10 jurnal terkait dengan perkembangan Preprocessing di NLP

| **No** | **Penulis/Sumber** | **Tahun** | **Judul Penelitian** | **Model** | **Pembahasan** | **Sumber** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Palomino, Aider | 2023 | Natural Language Processing: Recent Development and Applications | *NLP* | Pengaruh preprocessing pada akurasi analisis sentimen media sosial | https://www.mdpi.com/2076-3417/13/20/11395 |
| 2 | Camacho-Collados et al. | 2017 | On the Role of Text Preprocessing in Neural Network Architectures |  | Evaluasi tokenizing, lemmatizing, lowercasing, multiword grouping pada neural classifier  Dan tokenisasi sederhana cukup, namun teknik lain berpengaruh tergantung tugas | https://ar5iv.labs.arxiv.org/html/1707.01780 |
| 3 | ScienceDirect (Review) | 2024 | Recent advancements and challenges of NLP-based sentiment analysis |  | Tinjauan preprocessing pada analisis sentimen (tokenization, normalization, stemming, dsb.) dan Preprocessing penting untuk hasil analisis sentimen yang lebih baik | https://www.sciencedirect.com/science/article/pii/S2949719124000074 |
| 4 | Meystre et al. | 2015 | Recent Advances in Clinical Natural Language Processing in Support of Semantic Analysis |  | Preprocessing dalam NLP klinis (normalisasi, tokenisasi, dsb.) dan Preprocessing mendukung ekstraksi informasi medis lebih akurat | https://pmc.ncbi.nlm.nih.gov/articles/PMC4587060/ |
| 5 | Accenture (dalam Yroge) | 2023 | Global NLP Market and Preprocessing Trends | *Prepocesing NLP* | Proyeksi pasar NLP dan pentingnya preprocessing dan Preprocessing menjadi kunci pertumbuhan aplikasi NLP | https://www.linkedin.com/pulse/enhancing-nlp-accuracy-power-text-preprocessing-techniques-yroge |
| 6 | LinkedIn (Yroge) | 2024 | Future Trends and Innovations in Text Preprocessing |  | Inovasi preprocessing (deep learning, automation, integrasi NER, dsb.) dan Preprocessing makin otomatis dan kontekstual | https://www.linkedin.com/pulse/enhancing-nlp-accuracy-power-text-preprocessing-techniques-yroge |
| 7 | ScienceDirect (Review) | 2024 | Pre-processing methodologies in sentiment analysis |  | Studi berbagai teknik preprocessing untuk analisis sentimen dan Kombinasi teknik preprocessing menghasilkan hasil terbaik | https://www.sciencedirect.com/science/article/pii/S2949719124000074 |
| 8 | Palomino & Aider | 2017 | Preprocessing for Social Media Sentiment Analysis |  | Perbandingan preprocessing untuk pelatihan word embeddings  Preprocessing berbeda berdampak pada hasil embedding | https://ar5iv.labs.arxiv.org/html/1707.01780 |
| 9 | Camacho-Collados et al. | 2017 | Variability in Preprocessing Techniques and Their Effects |  | Variasi teknik preprocessing dan pengaruhnya pada model NLP dan Variasi teknik menghasilkan variabilitas performa model | https://ar5iv.labs.arxiv.org/html/1707.01780 |
| 10 | Meystre et al. | 2015 | Semantic Analysis Supported by Preprocessing in Clinical NLP |  | Variasi teknik preprocessing dan pengaruhnya pada model NLP dan Variasi teknik menghasilkan variabilitas performa model | https://ar5iv.labs.arxiv.org/html/1707.01780 |

# PERTEMUAN 4

## Cari 10 jurnal terkait dengan perkembangan Feature Extraction di NLP

| **No** | **Penulis/Sumber** | **Tahun** | **Judul Penelitian** | **Model** | **Pembahasan** | **Sumber** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Xiaobing Sun, Jiaxi Li, Wei Lu | 2023 | Unraveling Feature Extraction Mechanisms in Neural Networks | *LLM,NLP* | Analisis teoretis mekanisme ekstraksi fitur pada neural networks, efek fungsi aktivasi, self-attention, dg kesimpulan Multiplication-based models unggul di n-gram, ReLU bias fitur, insight untuk LLM | https://aclanthology.org/2023.emnlp-main.650/ |
| 2 | S. Palomino, A. Aider | 2023 | Natural Language Processing: Recent Development and Applications | NLP,BERT | Survei teknik NLP, termasuk ekstraksi fitur klasik dan deep learning dg kesimpulan Deep learning (Word2Vec, BERT) mengungguli teknik klasik seperti BoW, TF-IDF | https://www.mdpi.com/2076-3417/13/11/6438 |
| 3 | Xiaobing Sun, Jiaxi Li, Wei Lu | 2021 | Advances In Natural Language Processing: A Survey Of Techniques | NLP | Survei perkembangan teknik NLP, termasuk evolusi feature extraction dg kesimpulan Transformer dan contextual embedding merevolusi ekstraksi fitur | https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=5064608 |
| 4 | S. Palomino, A. Aider | 2023 | Preprocessing for Social Media Sentiment Analysis | *NLP* | Studi pengaruh preprocessing dan feature extraction pada data media sosial dg kesimpulan Preprocessing dan feature extraction meningkatkan performa klasifikasi sentimen | https://www.mdpi.com/2076-3417/13/11/6438 |
| 5 | S. Palomino, A. Aider | 2023 | Preprocessing Impact on Sentiment Classifier Performance | *NLP* | Pengaruh preprocessing dan feature extraction pada performa klasifikasi sentimen dg kesimpulan Kombinasi preprocessing dan feature extraction signifikan meningkatkan akurasi | https://www.mdpi.com/2076-3417/13/11/6438 |
| 6 | D. Supriyanto et al. | 2024 | Development of extraction features for Detecting Adolescent Personality with Machine Learning Algorithms | Naïve Bayes | Kombinasi TF-IDF + N-Gram Z untuk klasifikasi kepribadian remaja dg kesimpulan TF-IDF+N-Gram Z + Naïve Bayes capai akurasi hingga 98% | https://joiv.org/index.php/joiv/article/view/3091/0 |
| 7 | Chao et al. | 2024 | Research on Features Extraction and Classification for Images based on Transformer Learning | *CNN,ML* | Framework transformer-based untuk ekstraksi fitur dan klasifikasi pada data gambar dg kesimpulan Transformer-based extraction unggul dibanding CNN/ML klasik pada dataset gambar | https://proceedings.mlr.press/v245/chao24a.html |
| 8 | Y. Zhang et al. | 2023 | TwIdw-A Novel Method for Feature Extraction from Unstructured Text for Fake News Classification | NLP | Pengembangan metode TwIdw untuk ekstraksi fitur pada deteksi berita palsudg kesimpulan TwIdw efektif meningkatkan akurasi deteksi fake news | https://www.mdpi.com/2076-3417/13/11/6438 |
| 9 | S. Palomino, A. Aider | 2024 | Pre-processing methodologies in sentiment analysis | NLP | Studi berbagai teknik preprocessing dan feature extraction untuk analisis sentimen dg kesimpulan Kombinasi teknik menghasilkan hasil terbaik | https://www.mdpi.com/2076-3417/13/11/6438 |
| 10 | S. Palomino, A. Aider | 2023 | Integrating Feature Extraction and Optimal Selection to Combat Fake News | NLP | Dual-stage feature extraction dan seleksi fitur untuk deteksi fake news dg kesimpulan Seleksi fitur setelah ekstraksi meningkatkan akurasi deteksi fake news | https://journal.esrgroups.org/jes/article/view/6017 |

# PERTEMUAN 5

5.1 10 jurnal terkait dengan Information Retrieval

| **No** | **Penulis/Sumber** | **Tahun** | **Judul Penelitian** | **Model** | **Pembahasan** | **Sumber** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Hope Nabankema | 2024 | Evaluation of Natural Language Processing Techniques for Information Retrieval |  | Evaluasi teknik NLP (tokenisasi, NER, semantic parsing, word embeddings) untuk IR di berbagai domain dg kesimpulan NLP signifikan meningkatkan akurasi dan efisiensi IR, rekomendasi integrasi teknik | hybridhttps://doi.org/10.47941/ejikm.1752 |
| 2 | Chuang, Jackson & Jensen | 2012 | Topic Modeling for Enhanced Information Retrieval |  | Integrasi topic modeling dalam algoritma pencarian dengan kesimpulan Memperluas eksplorasi konten dan hasil pencarian tematis | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3400142/ |
| 3 | Luong, Pham & Manning | 2013 | Word Embeddings and Semantic Search for Information Retrieval |  | Penerapan embeddings untuk semantic search dan query expansion dengan kesimpulan Semantic search dengan embeddings meningkatkan relevansi hasil pencarian | https://aclanthology.org/D13-1170/ |
| 4 | Nivre et al. | 2020 | Dependency Parsing in Information Retrieval |  | Analisis struktur sintaksis untuk meningkatkan pemahaman query dan ranking dokumen dengan kesimpulan Dependency parsing memperbaiki pencocokan query-dokumen kompleks | https://aclanthology.org/2020.lrec-1.348/ |
| 5 | Le, Mikolov | 2014 | Distributed Representations of Sentences and Documents |  | Doc2vec untuk representasi dokumen dalam IR dengan kesimpulan Meningkatkan akurasi pencarian dokumen berbasis semantik | https://arxiv.org/abs/1405.4053 |
| 6 | Blei, D.M. | 2003 | Topic Modeling and Information Retrieval |  | Penggunaan topic modeling untuk identifikasi tema dokumen dan peningkatan IR dengan kesimpulan Topic modeling meningkatkan clustering dan relevansi hasil pencarian | https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf |
| 7 | Mikolov et al. | 2013 | Word Embeddings for Information Retrieval |  | Penggunaan word2vec/GloVe untuk representasi semantik kata pada IR dengan kesimpulan Word embeddings meningkatkan relevansi ranking dan query expansion | https://arxiv.org/abs/1301.3781 |
| 8 | Bast & Haussmann | 2015 | Semantic Parsing for Information Retrieval |  | Penggunaan semantic parsing untuk pemahaman query dan pencarian informasi pada data terstruktur/semi-terstruktur dengan kesimpulan Semantic parsing meningkatkan presisi dan recall IR | https://www.semanticscholar.org/paper/Parsing-as-Reduction-Bast-Haussmann/6e5e0e8b7b2c1b1b |
| 9 | Kim & Lee | 2020 | Meta-Analysis of NLP Techniques for Information Retrieval |  | Meta-analisis teknik NLP (tokenization, NER, semantic parsing) untuk peningkatan akurasi IR dengan kesimpulan Hybrid NLP methods (word embeddings + topic modeling) unggul dalam IR | https://carijournals.org/journals/index.php/EJIKM/article/view/1752 |
| 10 | MDPI (Special Issue Editors) | 2023 | Natural Language Processing and Information Retrieval (Special Issue) |  | Kumpulan makalah terbaru tentang teori, model, dan aplikasi NLP untuk IR dengan kesimpulan NLP memperluas aplikasi IR di berbagai domain dan meningkatkan semantic search | https://www.mdpi.com/journal/electronics/special\_issues/natural\_language\_information\_retrieval |

# PERTEMUAN 6

6.1 10 Jurnal terkait pemanfaatan Topic Model (LDA dan turunannya/related work)

| **No** | **Penulis/Sumber** | **Tahun** | **Judul Penelitian** | **Model** | **Pembahasan** | **Sumber** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Maarten Grootendorst et al. | 2025 | AI-powered topic modeling: comparing LDA and BERTopic |  | Perbandingan LDA (MALLET) dan BERTopic (BioBERT, UMAP, HDBSCAN) dengan integrasi AI (ChatGPT-4-Turbo) pada data medis dengan kesimpulan BERTopic lebih baik dalam koherensi semantik dan interpretabilitas; LDA butuh preprocessing manual; AI meningkatkan hasil | https://www.ebm-journal.org/journals/experimental-biology-and-medicine/articles/10.3389/ebm.2025.10389/full |
| 2 | Xuefeng Zhu et al. | 2024 | Investigating topic modeling techniques through evaluation of short texts |  | Evaluasi LDA dan NMF pada dataset teks pendek (SemEval 2016, BBC News) dengan clustering dan silhouette analysis  Model evaluasi baru mengungguli LDA dan NMF; penemuan topik baru | https://www.nature.com/articles/s41598-024-61738-4 |
| 3 | Maarten Grootendorst et al. | 2025 | AI-powered topic modeling: comparing LDA and BERTopic |  | Sama seperti jurnal SEBM, fokus pada analisis topik terkait risiko kardiovaskular opioid pada wanita  BERTopic menghasilkan cluster lebih kompak dan interpretasi otomatis dengan AI | https://pmc.ncbi.nlm.nih.gov/articles/PMC11906279/ |
| 4 | Abdullah Al Mamun et al. | 2024 | Evaluating the latest trends of Industry 4.0 based on LDA topic model |  | LDA diterapkan untuk mengidentifikasi pola tersembunyi dalam penelitian Industry 4.0  LDA efektif mengungkap tren riset Industry 4.0 | https://dl.acm.org/doi/10.1007/s11227-024-06247-x |
| 5 | S. S. S. R. K. Prasad et al. | 2025 | A Performance-Driven Exploration of Combining Topic Modeling and Deep Learning |  | Kombinasi LDA dan deep learning untuk peningkatan kualitas topic modeling  Kombinasi metode meningkatkan akurasi dan interpretabilitas topik | https://www.temjournal.com/content/141/TEMJournalFebruary2025\_511\_527.pdf |
| 6 | S. Wang et al. | 2024 | A topic modeling approach for analyzing and categorizing electronic health records |  | Topic modeling untuk analisis dan kategorisasi data rekam medis elektronik  Model topic membantu klasifikasi dan ekstraksi informasi medis | https://www.nature.com/articles/s41598-024-83743-3 |
| 7 | Md. Tanjim Hossain et al. | 2024 | Combining BERT with LDA: Improved Topic Modeling in Bengali Language |  | Integrasi BERT embeddings dengan LDA untuk topic modeling bahasa Bengali  Peningkatan kualitas topik dan interpretasi dibanding LDA tradisional | https://www.iaeng.org/IJCS/issues\_v52/issue\_2/IJCS\_52\_2\_11.pdf |
| 8 | Y. Zhang et al. | 2024 | Application of structural topic modeling in a literature review of air pollution research |  | Structural Topic Modeling (STM) diterapkan untuk review literatur polusi udara  STM efektif mengidentifikasi tema dan tren riset polusi udara | https://www.sciencedirect.com/science/article/abs/pii/S096969972400173X |
| 9 | Feliciaa  Muhammad Rizky Pribadi | 2024 | *Analisis Interaksi Pengguna Sosial Media Sekolah di Palembang Berdasarkan Topik dengan hLDA dan SVM* | SVM, LDA | Pada proses klasifikasi dataset dibagi menjadi 70% untuk training dan 30% untuk testing, dengan evaluasi berdasarkan F1-Score. Hasil terbaik diperoleh oleh SVM-SMOTE, dengan nilai F1-Score terbaik dari Dataset hLDA 3 Level (13 label), mencapai 95.68% dan nilai terendah dari Dataset hLDA 5 Level (8 label), mencapai 79.43%. Dataset yang memiliki lebih banyak topik memberikan hasil klasifikasi yang lebih baik. Berdasarkan jumlah like setiap topik Dataset hLDA 3 Level, yang paling diminati adalah topik 11 yang meliputi fasilitas sekolah, seragam murid, dan event hiburan. Informasi ini dapat membantu sekolah untuk mengembangkan lebih lanjut topik yang paling diminati serta meningkatkan topik yang kurang diminati. | https://jurnal.unprimdn.ac.id/index.php/JUTIKOMP/article/view/5536 |
| 10 | Springer, Cham | 2024 | Comparison of LDA, NMF and BERTopic Topic Modeling Techniques on Amazon Product Review Dataset: A Case Study | LDA,NMF,BERT | Dengan algoritma pemodelan topik, keluhan pengguna dapat dikelompokkan dan dibaca dalam kelompok. Dalam penelitian ini, LDA (Latent Dirichlet allocation), NMF (Non-Negative Matrix Factorization) dan algoritma BERTopic yang diuji pada kumpulan data ulasan produk Amazon dibandingkan. Menurut hasil yang diperoleh, semua 3 algoritma berhasil dan berguna. Algoritma BERTopic menghasilkan hasil yang lebih bermakna daripada algoritma lain sesuai dengan metrik perhitungan konsistensi. | https://link.springer.com/chapter/10.1007/978-3-031-53717-2\_3 |