

202110370311193  
Muhammad Bima Mauludin  
Prodi Informatika

**PENDEKATAN HIBRIDA PERINGKASAN TRANSKRIPSI  
VIDEO DENGAN KLASTERISASI HIERARCHICAL BERBASIS  
BERT DAN BART**

**TUGAS AKHIR**

Diajukan Untuk Memenuhi  
Persyaratan Guna Meraih Gelar Sarjana  
Informatika Universitas Muhammadiyah Malang



Muhammad Bima Mauludin

202110370311193

**Data Science**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MALANG**

**2025**

202110370311193  
Muhammad Bima Mauludin  
Prodi Informatika

**LEMBAR PERSETUJUAN**  
**PENDEKATAN HIBRIDA PERINGKASAN TRANSKRIPSI**  
**VIDEO DENGAN KLASTERISASI HIERARCHICAL BERBASIS**  
**BERT DAN BART**

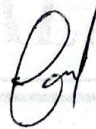
**TUGAS AKHIR**

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang

**Menyetujui,**

**Malang, 29 Desember 2025**

**Dosen Pembimbing 1**



**Christian Sri Kusuma Aditya, S.Kom., M.Kom.**

**NIP. 180327021991**

## LEMBAR PENGESAHAN

### Pendekatan Hibrida Peringkasan Transkripsi Video dengan Klusterisasi Hierarchical Berbasis BERT dan BART

### TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

**MUHAMMAD BIMA MAULUDIN**

**202110370311193**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji  
pada tanggal 17 April 2026

Menyetujui,

Dosen Pembimbing 1



Christian Sri Kusuma Aditya S.Kom.,

M.Kom

NIP. 180327021991PNS.

Dosen Pembimbing 2



NIP.

Dosen Penguji 1



Ir. Agus Eko Minarno S.Kom., M.Kom.,

IPM.

NIP. 10814100540PNS.

Dosen Penguji 2



Diah Risqiwati ST., MT.

NIP. 10814100545PNS.

Mengetahui,  
Ketua Jurusan Informatika



Ir. Agus Eko Minarno S.Kom., M.Kom., IPM.

NIP. 10814100540PNS.



## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**NAMA** : MUHAMMAD BIMA MAULUDIN  
**NIM** : 202110370311193  
**FAK./JUR.** : INFORMATIKA

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“Pendekatan Hibrida Peringkasan Transkripsi Video dengan Klasterisasi Hierarchical Berbasis BERT dan BART”** beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Malang, 20 Agustus 2025

Mengetahui,  
Dosen Pembimbing



Christian Sri Kusuma Aditya,  
S.Kom., M.Kom.

Yang membuat pernyataan



Muhammad Bima Mauludin

## ABSTRAK

Penggunaan video sebagai media penyampaian informasi dan edukasi terus mengalami peningkatan, namun durasi yang panjang serta struktur percakapan yang beragam sering menyulitkan proses pemahaman inti materi. Hal ini menjadi tantangan dalam pengembangan metode peringkasan otomatis, khususnya pada konten monolog, wawancara, podcast, dan short interview. Metode ekstraktif umumnya mampu mempertahankan informasi penting tetapi kurang koheren, sementara metode abstraktif cenderung lebih ringkas namun berisiko kehilangan konteks. Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan pendekatan hibrida yang mengombinasikan peringkasan ekstraktif berbasis representasi kalimat menggunakan BERT dan klasterisasi hierarchial, serta peringkasan abstraktif menggunakan model BART. Hasil evaluasi menunjukkan bahwa pendekatan hibrida memberikan performa yang kompetitif dan cenderung unggul pada konten dengan struktur percakapan kompleks, seperti wawancara dan podcast, dengan nilai rata-rata ROUGE-1 masing-masing sebesar 0.30 dan 0.32. Pendekatan ini juga mampu menghasilkan ringkasan yang lebih padat dan informatif dibandingkan metode tunggal, meskipun pada konten yang lebih singkat dan terstruktur seperti monolog dan short interview, perbedaannya tidak selalu dominan. Temuan ini menegaskan bahwa pendekatan hibrida memiliki potensi yang kuat sebagai solusi peringkasan transkripsi video yang adaptif terhadap berbagai karakteristik konten.

***Kata Kunci*** ~ Perangkum Video; Embedding BERT; Klastering Hierarchical

## LEMBAR PERSEMBAHAN

Segala puji dan syukur saya panjatkan ke hadirat Allah SWT atas rahmat, kekuatan, dan pertolongan-Nya yang senantiasa menyertai setiap langkah saya. Dengan segala keterbatasan dan perjuangan yang telah dilalui, akhirnya karya ini dapat diselesaikan. Dengan penuh kerendahan hati, karya ini saya persembahkan kepada:

1. **Keluarga saya tercinta**, yang tanpa lelah selalu hadir dalam setiap doa dan harapan saya. Di saat saya merasa lemah, ragu, dan hampir menyerah, kalian tetap menjadi tempat saya berpulang dan menguatkan. Terima kasih atas kasih sayang yang tulus, kesabaran yang tak terucap, serta dukungan yang tidak pernah berhenti mengalir. Segala pencapaian ini tidak akan berarti tanpa kalian di belakang saya.
2. **Diri saya sendiri**, yang telah bertahan melewati hari-hari berat, malam-malam panjang, dan rasa lelah yang sering kali tidak terlihat oleh siapa pun. Terima kasih karena tidak memilih untuk berhenti, meskipun jalan yang ditempuh terasa sulit. Karya ini menjadi saksi bahwa setiap air mata, usaha, dan doa yang dipanjatkan tidak pernah sia-sia.
3. **Bapak Christian Sri Kusuma Aditya, S.Kom., M.Kom.**, selaku dosen pembimbing, yang dengan kesabaran dan ketulusan telah membimbing saya di tengah berbagai keterbatasan dan kekurangan. Terima kasih atas arahan, perhatian, serta kepercayaan yang diberikan, sehingga saya mampu menyelesaikan penelitian ini dengan sebaik mungkin.
4. **Seluruh dosen Program Studi Informatika Universitas Muhammadiyah Malang**, yang telah memberikan ilmu, nilai, dan pembelajaran kehidupan selama masa perkuliahan. Setiap nasihat, teguran, dan pengajaran yang diberikan menjadi bekal berharga dalam membentuk cara berpikir dan sikap saya hingga hari ini.
5. **Teman-teman dalam grup Calon Pegawai NASA**, yang setia menemani perjalanan saya sejak semester pertama hingga akhir masa studi. Terima kasih telah tumbuh dan berjuang bersama, saling menguatkan di tengah tekanan, serta berbagi

cerita di saat lelah dan hampir menyerah. Kebersamaan ini akan selalu menjadi bagian berharga dalam perjalanan hidup saya.

6. **Almamater Universitas Muhammadiyah Malang**, tempat saya belajar tentang arti perjuangan, proses, dan tanggung jawab. Terima kasih telah menjadi rumah kedua yang menempa saya melalui berbagai pengalaman, hingga akhirnya saya dapat berdiri di titik ini.



## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penelitian ini disusun sebagai salah satu bentuk pemenuhan kewajiban akademik di Universitas Muhammadiyah Malang dengan judul “*Pendekatan Hibrida Peringkasan Transkripsi Video dengan Klasterisasi Hierarchical Berbasis BERT dan BART.*”

Penelitian ini dilatarbelakangi oleh permasalahan dalam pengolahan transkripsi video yang cenderung panjang dan tidak terstruktur. Untuk mengatasinya, penelitian ini mengusulkan pendekatan peringkasan hibrida yang mengombinasikan metode ekstraktif berbasis BERT dengan klasterisasi hierarchial serta metode abstraktif menggunakan model BART. Evaluasi kualitas ringkasan dilakukan menggunakan metrik ROUGE dengan membandingkannya terhadap ringkasan acuan.

Penulis mengucapkan terima kasih kepada Bapak **Christian Sri Kusuma Aditya, S.Kom., M.Kom.** selaku dosen pembimbing atas bimbingan dan arahnya, serta kepada keluarga dan seluruh pihak yang telah memberikan dukungan. Penulis menyadari penelitian ini masih memiliki keterbatasan dan terbuka terhadap kritik serta saran yang membangun demi penyempurnaan di masa mendatang.

Malang, 20 Agustus 2025



Muhammad Bima Mauludin

## DAFTAR ISI

<b>COVER .....</b>	<b>I</b>
<b>LEMBAR PERSETUJUAN .....</b>	<b>II</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>III</b>
<b>LEMBAR PERNYATAAN .....</b>	<b>IV</b>
<b>ABSTRAK .....</b>	<b>V</b>
<b>LEMBAR PERSEMBAHAN .....</b>	<b>VI</b>
<b>KATA PENGANTAR .....</b>	<b>VII</b>
<b>DAFTAR ISI .....</b>	<b>VIII</b>
<b>DAFTAR TABEL .....</b>	<b>X</b>
<b>DAFTAR GAMBAR .....</b>	<b>XI</b>
<b>DAFTAR RUMUS .....</b>	<b>XII</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1. LATAR BELAKANG .....	1
1.2. RUMUSAN MASALAH .....	3
1.3. TUJUAN PENELITIAN .....	3
1.4. BATASAN PENELITIAN .....	3
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>5</b>
2.1. PERINGKASAN TEKS (TEXT SUMMARIZATION) .....	5
2.2. TRANSKRIPSI VIDEO .....	5
2.3. MODEL BERT UNTUK REPRESENTASI KALIMAT .....	6
2.4. ALGORITMA K-MEANS CLUSTERING .....	7
2.5. METRIK EVALUASI ROUGE .....	8
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>12</b>

3.1.	RANCANGAN PENELITIAN .....	12
3.2.	SUMBER DAN AKUISISI DATA .....	12
3.3.	PROSEDUR PENELITIAN .....	15
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>21</b>
4.1.	HASIL TRANSKRIPSI .....	21
4.2.	HASIL RINGKASAN EKSTRAKTIF.....	22
4.3.	EVALUASI KUALITAS RINGKASAN MENGGUNAKAN ROUGE.....	34
4.4.	SARAN UNTUK PENELITIAN SELANJUTNYA.....	38
<b>BAB V KESIMPULAN.....</b>		<b>39</b>
<b>DAFTAR PUSTAKA.....</b>		<b>41</b>



## DAFTAR TABEL

Tabel 1. Detail Video yang Digunakan .....	13
Tabel 2. Menjelaskan komponen dari metrik ROUGE.....	20
Tabel 3. Hasil Transkripsi Video .....	21
Tabel 4. Pedoman Panjang Ringkasan Berdasarkan Jumlah Kalimat.....	23
Tabel 5. Salah Satu Contoh Representatif yang Dipilih dari Video Monolog .....	21
Tabel 6. Nilai Cluster dan Jumlah Klaster yang Dihasilkan per Video .....	27
Tabel 7. Jumlah Kata dan Kalimat dalam Hasil Ringkasan Ekstraktif .....	29
Tabel 8. ROUGE F1 Scores untuk Ringkasan Ekstraktif.....	25
Tabel 9. Perbandingan Rata-rata Skor ROUGE F1.....	26



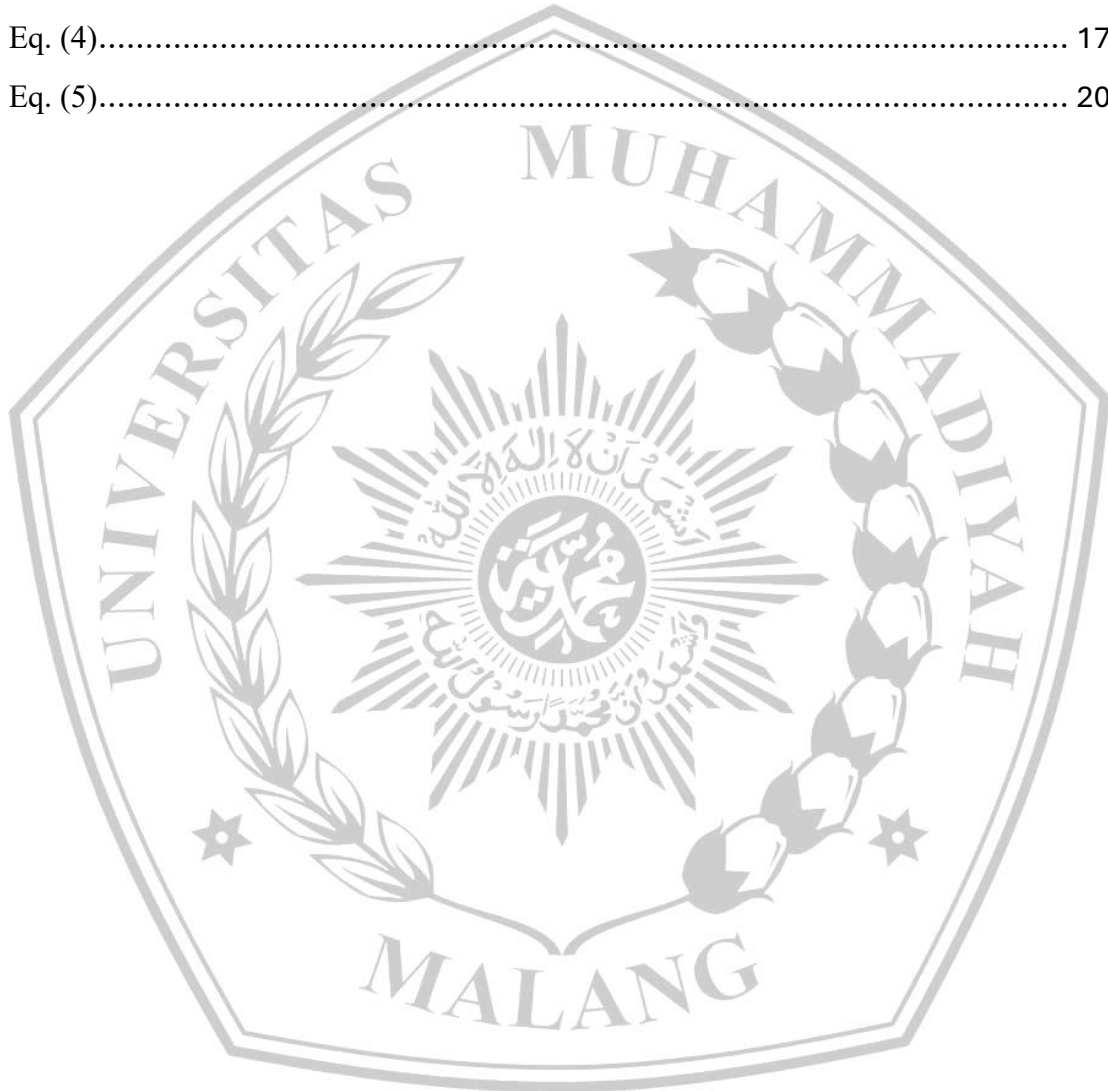
## DAFTAR GAMBAR

Gambar 1. Alur Penelitian untuk Ringkasan Otomatis Video HMPV .....	12
Gambar 2. Salah Satu Visualisasi Pengelompokan Hierachical untuk Monolog .....	24
Gambar 3. Salah Satu Visualisasi Pengelompokan Hierachical untuk Interview .....	25
Gambar 4. Salah Satu Visualisasi Pengelompokan Hierachical untuk Podcast.....	26
Gambar 5. Salah Satu Visualisasi Pengelompokan Hierachical untuk Short Interview .....	26



## DAFTAR RUMUS

Eq. (1).....	7
Eq. (2).....	8
Eq. (3).....	17
Eq. (4).....	17
Eq. (5).....	20



## DAFTAR PUSTAKA

- [1] J. P. Suharsono and D. Nurahman, "Pemanfaatan Youtube Sebagai Media Peningkatan Pelayanan Dan Informasi," *Ganaya J. Ilmu Sos. Dan Hum.*, vol. 7, no. 1, pp. 298–304, Mar. 2024, doi: 10.37329/ganaya.v7i1.3157.
- [2] K. Rangarajan, K. Begg, and B. Somani, "Online Digital Media: The Uptake of YouTube-based Digital Clinical Education (DCE)," *Am. J. Distance Educ.*, vol. 33, no. 2, pp. 142–150, Apr. 2019, doi: 10.1080/08923647.2019.1582308.
- [3] M. J. Barton, M. Okada, and M. Todorovic, "Podcasts in health education—Insights from a scoping review and survey," *Anat. Sci. Educ.*, vol. 18, no. 12, pp. 1388–1405, Dec. 2025, doi: 10.1002/ase.70037.
- [4] Neya Global and A. N. Kazanskaia, "Engaging Audiences in Public Speaking: Strategies for Effective Communication," *NEYA Glob. J. Non-Profit Stud.*, 2025, doi: 10.64357/neya-gjnps-eng-au-tp-tch-01.
- [5] M. S. Martín, J. Heras, and G. Mata, "Automatic Generation of Subtitles for Videos of the Government of La Rioja," in *Optimization and Learning*, vol. 1824, B. Dorronsoro, F. Chicano, G. Danoy, and E.-G. Talbi, Eds., in *Communications in Computer and Information Science*, vol. 1824, Cham: Springer Nature Switzerland, 2023, pp. 393–402. doi: 10.1007/978-3-031-34020-8\_30.
- [6] K. Lavanya, A. K. V. Dixit, and A. S. A., "Advanced Video Transcription And Summarization A Synergy of Langchain, Language Models, And VectorDB with Mozilla Deep Speech," in *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, Vellore, India: IEEE, Feb. 2024, pp. 1–9. doi: 10.1109/ic-ETITE58242.2024.10493791.
- [7] K. Song, C. Li, X. Wang, D. Yu, and F. Liu, "Towards Abstractive Grounded Summarization of Podcast Transcripts," 2022, *arXiv*. doi: 10.48550/ARXIV.2203.11425.
- [8] N. C. P. Prakash, A. P. Narasimhaiah, J. B. Nagaraj, P. K. Pareek, N. B. Maruthikumar, and R. I. Manjunath, "Implementation of NLP based automatic text summarization using spacy," *Int. J. Health Sci.*, pp. 7508–7521, Aug. 2022, doi: 10.53730/ijhs.v6nS5.10574.

- [9] N. Almohaimed and A. M. Azmi, "Abstractive text summarization: A comprehensive survey of techniques, systems, and challenges," *Comput. Sci. Rev.*, vol. 57, p. 100762, Aug. 2025, doi: 10.1016/j.cosrev.2025.100762.
- [10] D. S., S. N., J. Andrew, and M. Mazzara, "Unified extractive-abstractive summarization: a hybrid approach utilizing BERT and transformer models for enhanced document summarization," *PeerJ Comput. Sci.*, vol. 10, p. e2424, Nov. 2024, doi: 10.7717/peerj-cs.2424.
- [11] N. Shafiq, I. Hamid, M. Asif, Q. Nawaz, H. Aljuaid, and H. Ali, "Abstractive text summarization of low-resourced languages using deep learning," *PeerJ Comput. Sci.*, vol. 9, p. e1176, Jan. 2023, doi: 10.7717/peerj-cs.1176.
- [12] F. V. P. Samosir, H. Toba, and M. Ayub, "BESKlus : BERT Extractive Summarization with K-Means Clustering in Scientific Paper," *J. Tek. Inform. Dan Sist. Inf.*, vol. 8, no. 1, Apr. 2022, doi: 10.28932/jutisi.v8i1.4474.
- [13] N. Sadashiv, A. Krishna Maiya, G. Shivareddy, and A. Reddy, "Abstractive and extractive based YouTube transcript summarization: a hybrid approach," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 40, no. 3, p. 1439, Dec. 2025, doi: 10.11591/ijeecs.v40.i3.pp1439-1452.
- [14] Deki Riana, Widodo, and Murién Nugraheni, "PERINGKASAN TEKS BERBAHASA INDONESIA MENGGUNAKAN TEKNIK EKSTRAKSI DENGAN ALGORITMA LATENT SEMANTIC ANALYSIS (LSA) DENGAN VARIASI TF-IDF UNTUK PERINGKASAN SINGLE DOCUMENT," *PINTER J. Pendidik. Tek. Inform. Dan Komput.*, vol. 8, no. 1, pp. 95–101, Jun. 2024, doi: 10.21009/pinter.8.1.10.
- [15] I. Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand, and P. K. Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India: IEEE, Jan. 2021, pp. 1310–1317. doi: 10.1109/ICICT50816.2021.9358703.
- [16] B. Gana, H. Allende-Cid, S. Rüping, M. Becerra-Rozas, and J. Zamora, "A systematic review of long document summarization methods: Evaluation metrics and approaches," *Neurocomputing*, vol. 655, p. 131287, Nov. 2025, doi: 10.1016/j.neucom.2025.131287.

- [17] R. C. Belwal and A. Gupta, "Automatic text summarization techniques: A categorization, evolution and future scope," *Eng. Appl. Artif. Intell.*, vol. 157, p. 111216, Oct. 2025, doi: 10.1016/j.engappai.2025.111216.
- [18] S. Sharma, G. Aggarwal, and B. K. Rai, "A survey on the dataset, techniques, and evaluation metric used for abstractive text summarization," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 22, no. 3, p. 681, Jun. 2024, doi: 10.12928/telkomnika.v22i3.25512.
- [19] N. Giarelis, C. Mastrokostas, and N. Karacapilidis, "Abstractive vs. Extractive Summarization: An Experimental Review," *Appl. Sci.*, vol. 13, no. 13, p. 7620, Jun. 2023, doi: 10.3390/app13137620.
- [20] H. Guo, R. Pasunuru, and M. Bansal, "Soft Layer-Specific Multi-Task Summarization with Entailment and Question Generation," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 687–697. doi: 10.18653/v1/P18-1064.
- [21] Lviv Polytechnic National University, Knyaz Roman Str. 5 Lviv, 79013, Ukraine, N. I. Boyko, O. A. Tkachyk, and Lviv Polytechnic National University, Knyaz Roman Str. 5 Lviv, 79013; Ukraine, "Hierarchical clustering algorithm for dendrogram construction and cluster counting," *Inform. Math. Methods Simul.*, vol. 13, no. 1–2, pp. 5–15, Apr. 2023, doi: 10.15276/imms.v13.no1-2.5.
- [22] M. Lewis *et al.*, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online: Association for Computational Linguistics, 2020, pp. 7871–7880. doi: 10.18653/v1/2020.acl-main.703.
- [23] A. Nenkova and K. McKeown, "Automatic Summarization," *Found. Trends® Inf. Retr.*, vol. 5, no. 2–3, pp. 103–233, Jun. 2011, doi: 10.1561/1500000015.
- [24] A. Sobanjo-ter Meulen, A. V. Gutierrez, O. Ruiz, J. Eeuwijk, H. Vroling, and N. Kanesa-thasan, "The Burden of Human Metapneumovirus (hMPV) Disease in Older and High-Risk Adults in Developed Countries: A Systematic Literature Review," *Infect. Dis. Ther.*, vol. 14, no. 8, pp. 1917–1933, Aug. 2025, doi: 10.1007/s40121-025-01187-2.

202110370311193  
Muhammad Bima Mauludin  
Prodi Informatika

UNIVERSITAS  
MUHAMMADIYAH  
MALANG



## FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Muhammad Bima Mauludin  
NIM : 202110370311193  
Judul TA : PENDEKATAN HIBRIDA PERINGKASAN TRANSKRIPSI  
VIDEO DENGAN KLASTERISASI HIERARCHICAL  
BERBASIS BERT DAN BART

### Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	8 %
2.	Bab 2 – Daftar Pustaka	25 %	0 %
3.	Bab 3 – Analisis dan Perancangan	25 %	0 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	0 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	0 %
6.	Makalah Tugas Akhir	20%	13 %

\*) Hasil cek plagiarisme diisi oleh pemeriksa (staf TU)

\*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)

  
(.....)



Kampus I  
Jl. Bandung 1 Malang, Jawa Timur  
P. +62 341 551 253 (Hunting)  
F. +62 341 460 435

Kampus II  
Jl. Bendungan Sutarni No 188 Malang, Jawa Timur  
P. +62 341 551 149 (Hunting)  
F. +62 341 582 060

Kampus III  
Jl. Raya Tlogomas No. 248 Malang, Jawa Timur  
P. +62 341 464 318 (Hunting)  
F. +62 341 460 435  
E: webmaster@umm.ac.id