

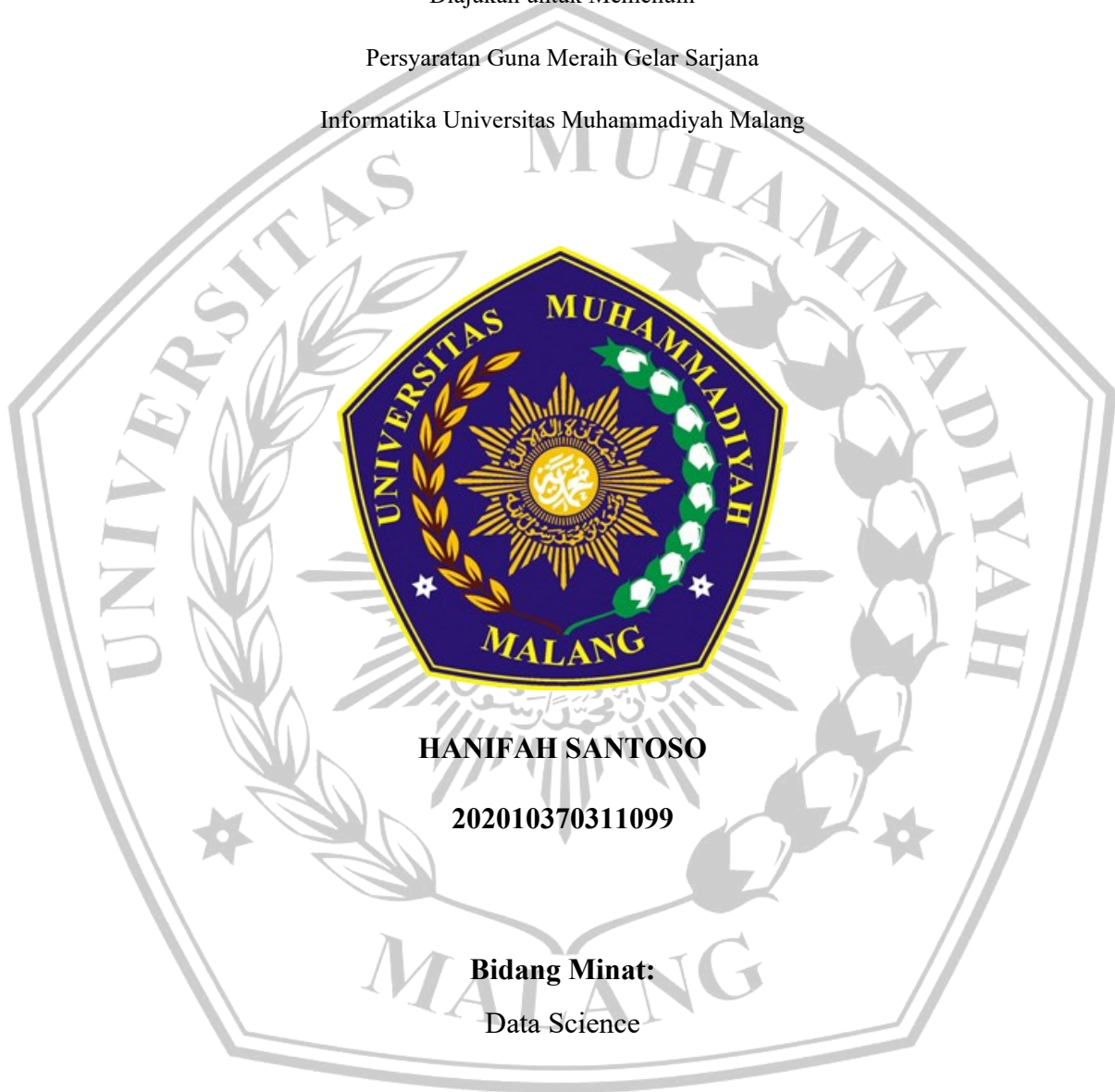
Analisis Sentimen Terkait Isu Penyebaran Nyamuk *Wolbachia* di Indonesia Pada Platform YouTube Menggunakan *Naïve Bayes*

Laporan Tugas Akhir

Diajukan untuk Memenuhi

Persyaratan Guna Meraih Gelar Sarjana

Informatika Universitas Muhammadiyah Malang



HANIFAH SANTOSO

202010370311099

Bidang Minat:

Data Science

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

2025

LEMBAR PERSETUJUAN

**Analisis Sentimen Terkait Isu Penyebaran Nyamuk Wolbachia di Indonesia Pada Platform
Youtube Menggunakan Nave Bayes**

TUGAS AKHIR

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

Menyetujui,

Malang, 20 Oktober 2025

Dosen Pembimbing 1

Dosen Pembimbing 2



Christian Sri Kusuma Aditya S.Kom.,

Didih Rizki Chandranegara S.kom.,

M.Kom

M.Kom

NIP. 180327021991PNS.

NIP. 180302101992PNS.

LEMBAR PENGESAHAN

Analisis Sentimen Terkait Isu Penyebaran Nyamuk Wolbachia di
Indonesia Pada Platform Youtube Menggunakan Nave Bayes

TUGAS AKHIR

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

Disusun Oleh :

Hanifah Santoso

202010370311099

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji
pada tanggal 20 Oktober 2025

Menyetujui,

Dosen Penguji 1



Nur Havatin S. ST., M.Kom

NIP. 10809070476PNS.

Dosen Penguji 2



Vinna Rahmayanti S S.Si., M.Si

NIP. 180306071990PNS.

Mengetahui,

Ketua Jurusan Informatika



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

NIP. 10814100540PNS.



LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : Hanifah Santoso

NIM : 202010370311099

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “Analisis Sentimen Terkait Isu Penyebaran Nyamuk Wolbachia di Indonesia Pada Platform Youtube Menggunakan Nave Bayes” 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.

Mengetahui,
Dosen Pembimbing



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

Malang, 20 Oktober 2025
Yang Membuat Pernyataan



Hanifah Santoso

ABSTRAK

Penyebaran nyamuk *Wolbachia* di Indonesia menjadi topik perdebatan di masyarakat dengan munculnya berbagai opini terkait efektivitas dan dampaknya. Namun, variasi sentimen belum dianalisis secara terstruktur dan terukur untuk polarisasi sentimen publik pada media YouTube. Oleh karena itu, penelitian ini bertujuan untuk menganalisis persepsi masyarakat terhadap program *Wolbachia* dengan menggunakan metode analisis sentimen berbasis *Multinomial Naïve Bayes* pada komentar YouTube. Metode penelitian mencakup *preprocessing* data teks, pembobotan kata menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*), serta penerapan tiga pendekatan klasifikasi: *Naïve Bayes Multinomial*, *Naïve Bayes Multinomial* dengan SMOTE, dan *Naïve Bayes Multinomial* dengan SMOTE serta *GridSearchCV* untuk optimasi *hyperparameter*. Hasil penelitian menunjukkan bahwa mayoritas sentimen masyarakat terhadap program *Wolbachia* bersifat negatif (67%), sementara 23% bersikap positif, dan 10% netral. Dari evaluasi model, pendekatan tanpa *balancing* data (Model 1) memberikan akurasi terbaik sebesar 70%, meskipun memiliki kelemahan dalam mengenali sentimen netral dan positif. Sedangkan pendekatan dengan SMOTE dan *GridSearchCV* (Model 3) menghasilkan akurasi sebesar 69% dengan keseimbangan yang lebih baik antara *precision* dan *recall*. Analisis *word cloud* menunjukkan bahwa opini negatif sering dikaitkan dengan teori konspirasi, ketidakpercayaan terhadap pemerintah, serta kekhawatiran terhadap dampak lingkungan. Kesimpulan dari penelitian menunjukkan bahwa persepsi masyarakat terhadap *Wolbachia* masih dipengaruhi oleh kurangnya edukasi dan transparansi informasi. Oleh karena itu, diperlukan peningkatan sosialisasi dan komunikasi yang lebih efektif guna mengurangi misinformasi serta meningkatkan penerimaan publik terhadap program ini.

Kata kunci: Analisis Sentimen, *Multinomial Naïve Bayes*, *Wolbachia*, SMOTE, *GridSearchCV*.

ABSTRACT

The spread of Wolbachia mosquitoes in Indonesia has become a topic of debate in society, with various opinions emerging regarding its effectiveness and impact. However, the variation in sentiment has not been analyzed in a structured and measurable way to determine the polarization of public sentiment on YouTube. Therefore, this study aims to analyze public perception of the Wolbachia program using a Multinomial Naïve Bayes sentiment analysis method on YouTube comments. The research method includes text data preprocessing, word weighting using TF-IDF (Term Frequency-Inverse Document Frequency), and the application of three classification approaches: Multinomial Naïve Bayes, Multinomial Naïve Bayes with SMOTE, and Multinomial Naïve Bayes with SMOTE and GridSearchCV for hyperparameter optimization. The results show that the majority of public sentiment towards the Wolbachia program is negative (67%), while 23% is positive and 10% is neutral. From the model evaluation, the approach without data balancing (Model 1) provides the best accuracy of 70%, although it has weaknesses in recognizing neutral and positive sentiments. Meanwhile, the approach using SMOTE and GridSearchCV (Model 3) produced an accuracy of 69% with a better balance between precision and recall. Word cloud analysis shows that negative opinions are often associated with conspiracy theories, distrust of the government, and concerns about environmental impacts. The conclusion of the study shows that public perception of Wolbachia is still influenced by a lack of education and transparency of information. Therefore, there is a need for more effective socialization and communication to reduce misinformation and increase public acceptance of this program.

Keywords: *Sentiment Analysis, Multinomial Naïve Bayes, Wolbachia, SMOTE, GridSearchCV.*

KATA PENGANTAR

Syukur Alhamdulillah penulis panjatkan ke hadirat Allah Ta'ala yang telah melimpahkan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul:

“Analisis Sentimen Terkait Isu Penyebaran Nyamuk *Wolbachia* di Indonesia Pada Platform YouTube Menggunakan *Naïve Bayes*”

Shalawat serta salam senantiasa tercurah kepada Allah Ta'ala, teladan bagi seluruh umat manusia yang telah mengajarkan nilai-nilai kebaikan dan ilmu pengetahuan sebagai cahaya dalam kehidupan.

Penelitian ini bertujuan untuk menganalisis sentiment masyarakat terhadap program *Wolbachia* yang dikembangkan sebagai strategi pengendalian Demam Berdarah *Dengue* (DBD) di Indonesia. Metode yang digunakan dalam penelitian adalah *Multinomial Naïve Bayes* yang merupakan salah satu pendekatan dalam bidang analisis sentimen berbasis *machine learning*. Hasil dari penelitian diharapkan dapat memberikan wawasan yang lebih jelas mengenai persepsi masyarakat terhadap program *Wolbachia*, serta menjadi referensi strategis bagi pemerintah dan pemangku kebijakan dalam merancang kebijakan yang lebih adaptif dan berbasis data.

Penulis menyadari sepenuhnya bahwa dalam penyusunan tugas akhir ini masih terdapat kekurangan dan keterbatasan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari berbagai pihak guna penyempurnaan di masa mendatang.

Akhir kata, semoga tugas akhir ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan serta menjadi kontribusi kecil dalam kemajuan akademik dan penelitian di bidang analisis sentimen dan pengolahan data tekstual. Semoga Allah SWT senantiasa memberikan keberkahan atas ilmu yang diperoleh dan menjadikannya bermanfaat bagi diri sendiri maupun orang lain.

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	iv
ABSTRAK	v
LEMBAR PERSEMBAHAN	vii
KATA PENGANTAR	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
Daftar Lampiran	xiv
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah	5
BAB II	6
TINJAUAN PUSTAKA	6
2.1 State of The Art (SOTA)	6
2.2 Machine Learning	9
2.3 Analisis Sentimen	9
2.4 Data Preprocessing	9
2.5 Kappa Statistik	10
2.6 TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)	12
2.7 <i>Multinomial Naive Bayes</i>	13
2.8 Confusion Matrix	15
BAB III	17
METODOLOGI PENELITIAN	17
3.1 Pengumpulan Data	17
3.2 Pelabelan Data	18
3.3 Preprocessing Data	19
3.4 TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)	20

3.5	Naïve Bayes Multinomial	21
3.6	Evaluasi Model	22
BAB IV		23
HASIL DAN PEMBAHASAN		23
4.1	Pengumpulan Data	23
4.2	Pelabelan Data	25
4.3	<i>Preprocessing Data</i>	28
4.4	Hasil TF-IDF	36
4.5	Implementasi Naive Bayes	38
4.5.1	Persiapan Data	39
4.5.2	Penerapan Model <i>Multinomial Naïve Bayes</i>	39
4.6	Pengujian dan Evaluasi Model	44
4.6.1	Pengujian Model	44
4.6.2	Evaluasi Model	49
4.6.3	Analisis Persepsi Masyarakat Terhadap Program <i>Wolbachia</i>	50
BAB V		53
PENUTUP		53
5.1	Kesimpulan	53
5.2	Saran	54
DAFTAR PUSTAKA		55
LAMPIRAN		58

DAFTAR GAMBAR

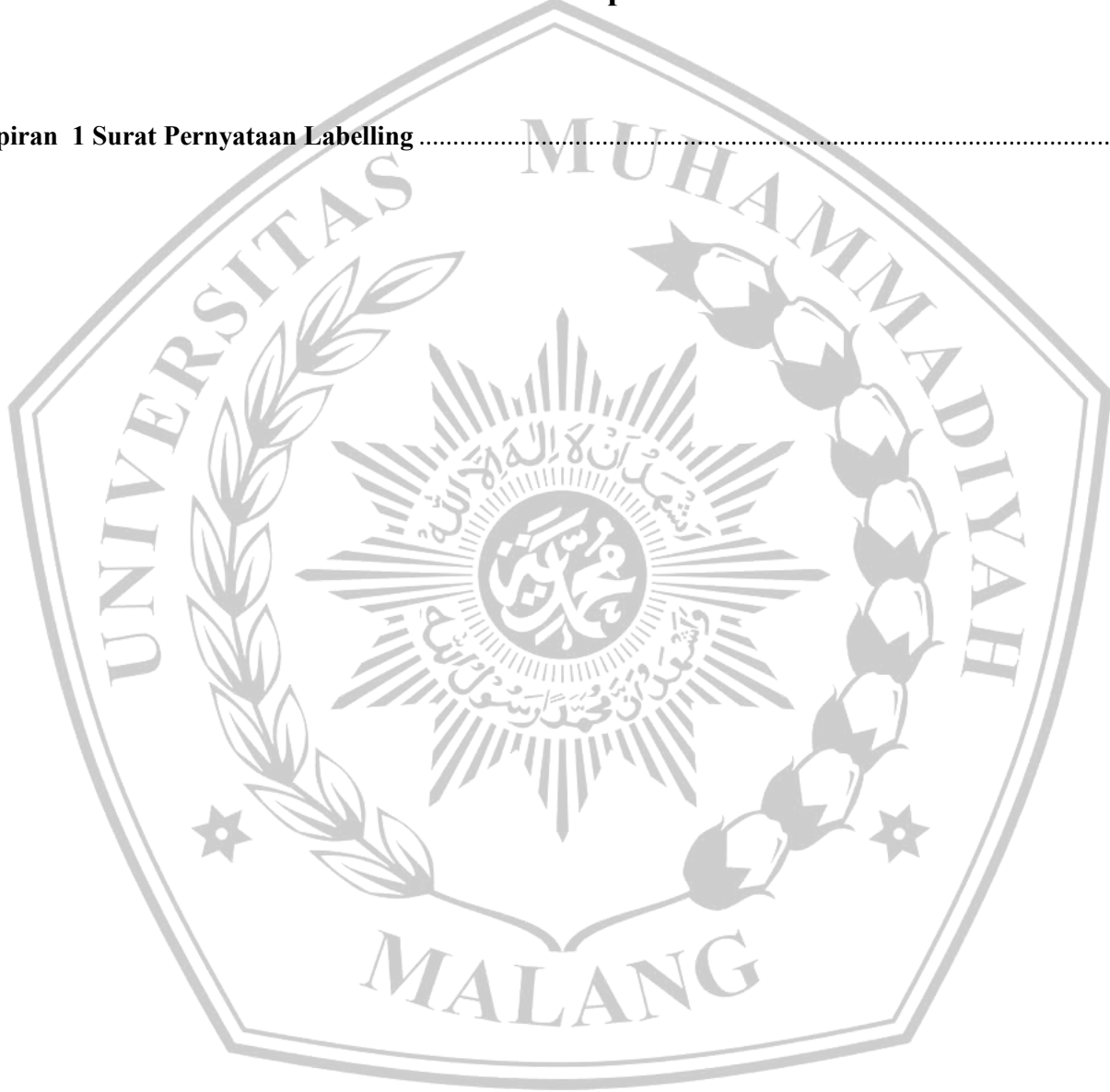
Gambar 3. 1 Rancangan Metode Penelitian.....	17
Gambar 4.1 Google Cloud.....	23
Gambar 4.2 Hasil Retrieval.....	25
Gambar 4.3 Data Labelling.....	26
Gambar 4. 4 Clean Text.....	33
Gambar 4. 5 Worcloud Positif.....	34
Gambar 4. 6 Barchart Positif.....	34
Gambar 4. 7 Wordcloud Netral.....	35
Gambar 4. 8 Barchart Netral.....	35
Gambar 4. 9 Wordcloud Negatif.....	36
Gambar 4. 10 Barchart Negatif.....	36
Gambar 4. 11 Confusion Matrix A.....	45
Gambar 4. 12 Confusion Matrix B.....	47
Gambar 4. 13 Confusion Matrix C.....	48

DAFTAR TABEL

Tabel 2. 1 State of The Art (SOTA).....	6
Tabel 2. 2 Penjelasan Teknik Preprocessing	10
Tabel 2. 3 Variasi Antar Penilai	11
Tabel 2. 4 Interpretasi Kappa.....	12
Tabel 2. 5 Confusion Matrix Multiclass.....	15
Tabel 3. 1 Keterangan Atribut.....	18
Tabel 3. 2 Tahap Preprocessing.....	20
Tabel 4. 1 Hasil Pelabelan Penilai	26
Tabel 4. 2 Hasil Variasi Label Antar Penilai.....	26
Tabel 4. 3 Parameter Pembobotan.....	37
Tabel 4. 4 Hasil TfidfVectorizer	38
Tabel 4.5 Hasil GridSearchCV	44
Tabel 4. 6 Classification Report A.....	44
Tabel 4. 7 Classification Report B	46
Tabel 4. 8 Classification Report C.....	48
Tabel 4. 9 Hasil Pengujian	49

Daftar Lampiran

Lampiran 1 Surat Pernyataan Labelling	58
---	----



DAFTAR PUSTAKA

- [1] D. Buchori, A. Mawan, I. Nurhayati, A. Aryati, H. Kusananto, dan U. K. Hadi, "Risk Assessment on the Release of Wolbachia-Infected *Aedes aegypti* in Yogyakarta, Indonesia," *Insects*, vol. 13, no. 10, 2022, doi: 10.3390/insects13100924.
- [2] T. S. Nadia, "Demam berdarah masih mengintai," *Mediakom*, no. April, 2024.
- [3] J. Venegas, S. Vaselek, dan M. F. Yasnot, "Editorial: Latest advances in the biological control of vectors of human tropical diseases," *Front. Trop. Dis.*, vol. 5, no. July, hal. 1–2, 2024, doi: 10.3389/fitd.2024.1430944.
- [4] C. P. Simmons *et al.*, "Successful introgression of wMel Wolbachia into *Aedes aegypti* populations in Fiji, Vanuatu and Kiribati," *PLoS Negl. Trop. Dis.*, vol. 18, no. 3, hal. 1–24, 2024, doi: 10.1371/journal.pntd.0012022.
- [5] F. Komariah, "Kemenkes: Pelepasan Nyamuk Wolbachia Masih 20 Persen," Radio Republik Indonesia. Diakses: 8 Oktober 2024. [Daring]. Tersedia pada: <https://rri.co.id/nasional/619286/kemenkes-pelepasan-nyamuk-wolbachia-masih-20-persen>
- [6] A. S. Cahyono, "Pengaruh media sosial terhadap perubahan sosial masyarakat di Indonesia," *J. Ilmu Sos. Ilmu Polit. diterbitkan oleh Fak. Ilmu Sos. Polit. Univ. Tulungagung*, vol. 9, no. 1, 2016.
- [7] cindy mutia Annur, "Pengguna YouTube di Indonesia Peringkat Keempat Terbanyak di Dunia pada Awal 2023," *Databoks*, 2023.
- [8] R. Novendri, A. S. Callista, D. N. Pratama, dan C. E. Puspita, "Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes," *Bull. Comput. Sci. Electr. Eng.*, vol. 1, no. 1, 2020, doi: 10.25008/bcsee.v1i1.5.
- [9] M. Birjali, M. Kasri, dan A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Syst.*, vol. 226, 2021, doi: 10.1016/j.knosys.2021.107134.
- [10] N. Hijriani dan E. Ermatita, "Perbandingan Naive Bayes Classifier dan Support Vector Machine untuk Analisis Sentimen terhadap Penyebaran Nyamuk Wolbachia di Indonesia," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 11, hal. 391–403, 2024, doi: 10.52436/1.jpti.499.

- [11] S. Styawati, N. Hendrastuty, dan A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, 2021, doi: 10.30591/jpit.v6i3.2870.
- [12] S. Mulyani dan R. Novita, "Implementation of The Naive Bayes Classifier Algorithm for Classification of Community Sentiment About Depression on Youtube," *J. Tek. Inform.*, vol. 3, no. 5, 2022, doi: 10.20884/1.jutif.2022.3.5.374.
- [13] R. Singh dan A. Tiwari, "YOUTUBE COMMENTS SENTIMENT ANALYSIS Analysis and Simulation of COVID-19 View project YOUTUBE COMMENTS SENTIMENT ANALYSIS," *Int. J. Sci. Res. Eng. Manag. (IJSREM)*, 2021.
- [14] N. Rezki, S. A. Thamrin, dan S. Siswanto, "SENTIMENT ANALYSIS OF MERDEKA BELAJAR KAMPUS MERDEKA POLICY USING SUPPORT VECTOR MACHINE WITH WORD2VEC," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 1, 2023, doi: 10.30598/barekengvol17iss1pp0481-0486.
- [15] A. Bulu, E. Umar, dan P. M. Ate, "Analisis Sentimen Terhadap Sistem Informasi Akademik STIMIKOM Stella Maris Sumba Menggunakan Algoritma Naïve Bayes," *J. Sist. Inf. Dan Inform.*, vol. 1, no. 2, 2023.
- [16] B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, vol. 9, no. 1, hal. 381–386, 2020, doi: 10.21275/art20203995.
- [17] T. Ahmad dan M. N. Aziz, "Data preprocessing and feature selection for machine learning intrusion detection systems," *ICIC Express Lett.*, vol. 13, no. 2, hal. 93–101, 2019, doi: 10.24507/icicel.13.02.93.
- [18] S. B. Kotsiantis dan D. Kanellopoulos, "Data preprocessing for supervised learning," *Int. J. ...*, vol. 1, no. 2, hal. 1–7, 2006, doi: 10.1080/02331931003692557.
- [19] Z. P. Putra dan A. Nugroho, "Pebandingan Performa Naïve Bayes dan KNN pada Klasifikasi Teks Sentimen Jasa Ekspedisi," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 6, no. 3, 2021, doi: 10.31328/jointecs.v6i3.2635.
- [20] A. Sabrani, I. G. W. Wedashwara W., dan F. Bimantoro, "Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia," *J. Teknol. Informasi, Komputer, dan Apl.*

(*JTIKA*), vol. 2, no. 1, hal. 89–100, 2020, doi: 10.29303/jtika.v2i1.87.

- [21] I. Amal dan Jayanta, “Perbandingan Pelabelan Otomatis Dan Manual Untuk Analisis Sentimen Terhadap Kenaikan Harga BBM Pertamina Pada Twitter Menggunakan Algoritma Support Vector Machine,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 4, no. 2, hal. 473–487, 2023, [Daring]. Tersedia pada: <https://conference.upnvj.ac.id/index.php/senamika/article/view/2562>
- [22] D. Musfiroh, U. Khaira, P. E. P. Utomo, dan T. Suratno, “Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, hal. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.
- [23] Z. Nazarudin, I. Muhimmah, dan I. Fidianingsih, “Segmentasi Citra untuk Menentukan Skor Kerusakan Hati secara Histologi,” *Semin. Nas. Inform. Medis*, hal. 15, 2017.
- [24] K. Adi, N. #1, dan D. Sebastian, “Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity,” *J. Tek. Inform. dan Sist. Inf.*, vol. 4, hal. 2443–2229, 2018, [Daring]. Tersedia pada: <http://dx.doi.org/10.28932/jutisi.v4i3.862>
- [25] Y. A. Putri Gabriella, “Optimasi Penerimaan Siswa Baru Dengan Penerapan Algoritma Text Mining Dan Tf-Idf,” *J. Comput. Informatics Res.*, vol. 2, no. 3, hal. 110–117, 2023, doi: 10.47065/comforch.v2i3.941.
- [26] Dicky Satria Mahendra, Basuki Rahmat, dan Retno Mumpuni, “Implementasi Metode Multinomial Naive Bayes dalam Klasifikasi Judul Berita Clickbait,” *Neptunus J. Ilmu Komput. Dan Teknol. Inf.*, vol. 2, no. 3, hal. 303–316, 2024, doi: 10.61132/neptunus.v2i3.249.



FAKULTAS TEKNIK

INFORMATIKA

informatika.umm.ac.id | informatika@umm.ac.id

UNIVERSITAS
MUHAMMADIYAH
MALANG



FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Hanifah Santoso

NIM : 202010370311099

Judul TA : Analisis Sentimen Terkait Isu Penyebaran Nyamuk Wolbachia di Indonesia Pada Platform YouTube Menggunakan Naïve Bayes



Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	0%
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 %	2%
6.	Makalah Tugas Akhir	20%	6%

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

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

Mengetahui,

Pemeriksa (Staff TU)

(.....
denf.....)



Kampus 1
Jl. Bendung 1 Malang, Jawa Timur
P. +62 341 551 253 (Pusat)
F. +62 341 460 430

Kampus 2
Jl. Bendungan Satria No. 153 Malang, Jawa Timur
P. +62 341 531 149 (Pusat)
F. +62 341 582 060

Kampus 3
Jl. Raya Tlogomas No. 246 Malang, Jawa Timur
P. +62 341 254 319 (Pusat)
F. +62 341 400 430
E. webmaster@umm.ac.id