

**PERBANDINGAN *NAIVE BAYES CLASSIFIER* DAN
SUPPORT VECTOR MACHINE UNTUK ANALISIS
SENTIMEN TERHADAP PENYEBARAN
NYAMUK *WOLBACHIA* DI INDONESIA**

TESIS

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Magister**



**OLEH :
NURUL HIJRIANI
09012682327009**

**MAGISTER ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA**

2024

**PERBANDINGAN *NAIVE BAYES CLASSIFIER* DAN
SUPPORT VECTOR MACHINE UNTUK ANALISIS
SENTIMEN TERHADAP PENYEBARAN
NYAMUK *WOLBACHIA* DI INDONESIA**

TESIS

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Magister**



**OLEH :
NURUL HIJRANI
09012682327009**

**MAGISTER ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2024**

HALAMAN PENGESAHAN

**PERBANDINGAN *NAIVE BAYES CLASSIFIER* DAN
SUPPORT VECTOR MACHINE UNTUK ANALISIS
SENTIMEN TERHADAP PENYEBARAN
NYAMUK *WOLBACHIA* DI INDONESIA**

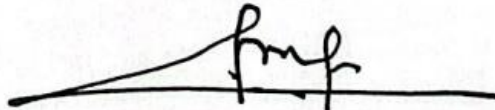
TESIS

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Magister

**OLEH:
NURUL HIJRIANI
09012682327009**

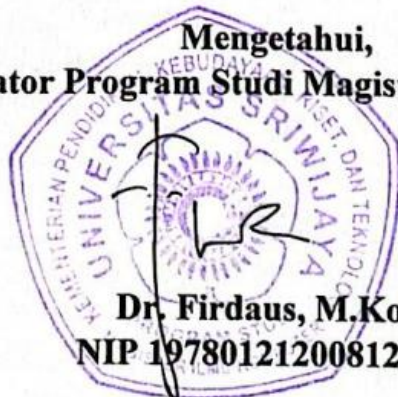
Palembang, 30 Desember 2024

Pembimbing



**Prof. Dr. Ermatita, M.Kom
NIP 196709132006042001**

**Mengetahui,
Koordinator Program Studi Magister Ilmu Komputer**



**Dr. Firdaus, M.Kom
NIP 197801212008121003**

Halaman Persetujuan

Pada hari Senin, tanggal 30 Desember 2024 telah dilaksanakan ujian sidang tesis oleh Magister Ilmu Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Nurul Hijriani
NIM : 09012682327009
Judul : Perbandingan *Naive Bayes Classifier* dan *Support Vector Machine* untuk Analisis Sentimen Terhadap Penyebaran Nyamuk *Wolbachia* di Indonesia

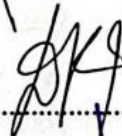
1. Ketua Penguji

Dr. Firdaus, M.Kom
NIP 197801212008121003



2. Penguji I

Dian Palupi Rini, M.Kom., Ph.D
NIP 197802232006042002



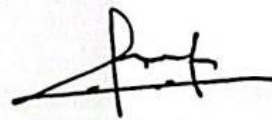
3. Penguji II

Dr. Abdiansah, S.Kom., M.CS
NIP 198410012009121005

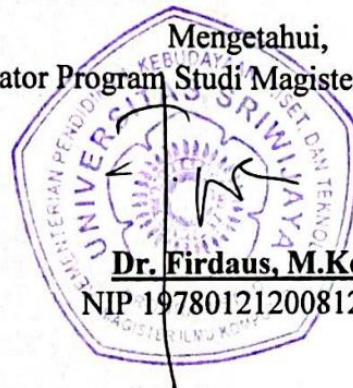


4. Pembimbing

Prof. Dr. Ermatita, M.Kom
NIP 196709132006042001



Mengetahui,
Koordinator Program Studi Magister Ilmu Komputer



Dr. Firdaus, M.Kom
NIP 197801212008121003

LEMBAR PERNYATAAN

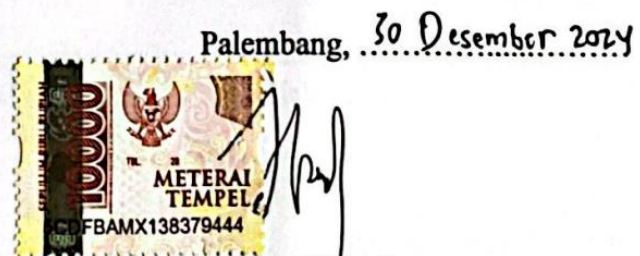
Yang bertanda tangan di bawah ini :

Nama : Nurul Hijriani
NIM : 09012682327009
Program Studi : Magister Ilmu Komputer
Judul Tesis : Perbandingan *Naive Bayes Classifier* dan *Support Vector Machine* untuk Analisis Sentimen Terhadap Penyebaran Nyamuk *Wolbachia* di Indonesia

Hasil Pengecekan Software iThenticate/Turnitin : 11 %

Menyatakan bahwa laporan tesis saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan tesis ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Nurul Hijriani

NIM 09012682327009

KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga saya dapat menyelesaikan tesis dengan judul "Perbandingan Naive Bayes Classifier dan Support Vector Machine untuk Analisis Sentimen Terhadap Nyamuk Wolbachia di Indonesia."

Penyusunan tesis ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, saya ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Ibu Prof. Dr. Ermatita, M.Kom, selaku pembimbing yang telah memberikan bimbingan, saran, dan motivasi dalam penyusunan tesis ini.
2. Bapak Hadipurnawan Satria, Ph.D. selaku pembimbing akademik yang telah memberikan bimbingan selama saya berkuliah.
3. Bapak Dr. Firdaus, M.Kom selaku Ketua Program Studi Magister Ilmu Komputer Universitas Sriwijaya
4. Keluarga saya, terutama anakku Hasbi, suamiku Iman, Nyai Maya, saudara-saudara ku, yang selalu kebersamai dan memberikan doa, dukungan moril, dan materiil.
5. Almarhum papa Suandi, yang selalu menjadi inspirasi dan sumber kekuatan dalam hidup saya.
6. Teman-teman dan rekan-rekan di Program Studi Magister Ilmu Komputer Universitas Sriwijaya yang telah memberikan bantuan dan dukungan.

Saya menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat saya harapkan untuk perbaikan di masa yang akan datang.

Akhir kata, semoga tesis ini dapat memberikan manfaat bagi semua pihak yang membutuhkan dan dapat menjadi referensi untuk penelitian lebih lanjut.

Palembang, 30 Desember 2024

Nurul Hijriani
NIM. 09012682327009

COMPARISON OF NAIVE BAYES CLASSIFIER AND SUPPORT VECTOR MACHINE FOR SENTIMENT ANALYSIS ON THE RELEASE OF WOLBACHIA MOSQUITOES IN INDONESIA

Nurul Hijriani (09012682327009)

Master's Department of Computer Science, Faculty of Computer Science, Sriwijaya University
Email: 09012682327009@student.unsri.ac.id

Abstract

*Dengue Hemorrhagic Fever (DHF) remains an endemic disease and a significant public health concern in Indonesia. One innovative solution implemented is the release of *Aedes aegypti* mosquitoes infected with Wolbachia bacteria to reduce dengue virus transmission. This program has received mixed responses from the public on social media platforms such as Instagram, which can influence its success. This study aims to analyze public sentiment regarding the program using Naive Bayes Classifier (NBC) and Support Vector Machine (SVM). A total of 3,619 Instagram comments were collected through scraping, processed using text preprocessing with N-Gram and TF-IDF methods for feature extraction. The Synthetic Minority Oversampling Technique (SMOTE) was applied to address data imbalance, and K-Fold Cross Validation with $k=10$ was used for model evaluation. Results indicate that SVM achieved higher accuracy at 84.6% compared to NBC at 80.3%, while NBC demonstrated superior recall for positive sentiment at 83.7%, making it effective in identifying public support. SVM, on the other hand, excelled in precision, reaching 80.7%, making it better at avoiding false positives. These findings suggest that NBC is more suitable for identifying support, while SVM is more reliable for generating precise predictions. This study provides valuable insights for government and stakeholders to develop better data-driven health communication strategies, supporting the success of public health programs.*

Keywords: *dengue fever, Instagram, Naive Bayes Classifier, sentiment analysis, Support Vector Machine, Wolbachia*

PERBANDINGAN *NAIVE BAYES CLASSIFIER* DAN *SUPPORT VECTOR MACHINE* UNTUK ANALISIS SENTIMEN TERHADAP PENYEBARAN NYAMUK *WOLBACHIA* DI INDONESIA

Nurul Hijriani (09012682327009)

Jurusan Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email: 09012682327009@student.unsri.ac.id

Abstrak

Demam Berdarah Dengue (DBD) masih menjadi penyakit endemik dan masalah kesehatan masyarakat yang signifikan di Indonesia. Salah satu solusi inovatif yang diterapkan adalah pelepasan nyamuk *Aedes aegypti* yang terinfeksi bakteri *Wolbachia* untuk mengurangi transmisi virus *dengue*. Program ini mendapatkan respons beragam dari masyarakat di platform media sosial seperti Instagram, yang dapat memengaruhi keberhasilannya. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program tersebut menggunakan metode *Naive Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)*. Sebanyak 3.619 komentar Instagram dikumpulkan melalui scraping, diproses menggunakan text preprocessing dengan metode N-Gram dan TF-IDF untuk ekstraksi fitur. Untuk mengatasi ketidakseimbangan data, *Synthetic Minority Oversampling Technique (SMOTE)* diterapkan, dan validasi dilakukan menggunakan *K-Fold Cross Validation* dengan $k=10$. Hasil menunjukkan bahwa SVM mencapai *accuracy* tertinggi sebesar 84.6% dibandingkan NBC yang mencapai 80.3%, sementara NBC memiliki *recall* lebih tinggi untuk sentimen positif sebesar 83.7%, sehingga lebih efektif untuk mengidentifikasi dukungan publik. Di sisi lain, SVM unggul dalam *precision*, mencapai 80.7%, yang membuatnya lebih andal dalam menghindari kesalahan prediksi sentimen positif. Temuan ini menyarankan bahwa NBC lebih cocok untuk mengidentifikasi dukungan, sementara SVM lebih handal untuk menghasilkan prediksi yang presisi. Penelitian ini memberikan wawasan praktis bagi pemerintah dan pemangku kepentingan untuk mengembangkan strategi komunikasi kesehatan berbasis data yang lebih baik, mendukung keberhasilan program kesehatan masyarakat.

Kata Kunci: *Analisis Sentiment, Demam Berdarah Dengue, Instagram, Naive Bayes Classifier, Support Vector Machine, Wolbachia*

DAFTAR ISI

Halaman Judul	ii
Halaman Pengesahan	iii
Halaman Persetujuan	iv
Lembar Pernyataan	v
KATA PENGANTAR	vi
<i>Abstract</i>	vii
Abstrak	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
DAFTAR ISTILAH	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA	7
2.1 Analisis Sentimen.....	7
2.2 Sosial Media Instagram	7
2.3 Nyamuk Wolbachia	8
2.4 <i>Valance Aware Dictionary and Sentiment Reasoner</i>	8
2.5 <i>K-Fold Cross Validation</i>	9
2.6 <i>Syntetic Minority Oversampling Technique</i>	9
2.7 <i>Text Preprocessing</i>	10
2.8 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	11
2.9 <i>N-Gram</i>	12
2.10 <i>Naive Bayes Classifier (NBC)</i>	12
2.11 <i>Support Vector Machine (SVM)</i>	15
2.12 <i>Confusion Matrix</i>	17
2.13 Penelitian Sejenis.....	18
BAB III METODOLOGI PENELITIAN	23
3.1 Alur Penelitian.....	23
3.2 Pengumpulan Data.....	23
3.2.1 Tipe Data.....	23
3.2.2 Teknik Pengumpulan Data	24

3.2.3 Penyimpanan Data.....	24
3.2.4 Pembagian Data ke Dalam Kelas	25
3.2.5 Kendala yang Dihadapi	25
3.3 <i>Text Preprocessing</i>	25
3.3.1 Penerapan SMOTE.....	28
3.3.2 Penerapan TF-IDF	29
3.3.3 Penerapan N-Gram.....	30
3.4 Klasifikasi	31
3.5 Evaluasi	35
3.5.1 Evaluasi NBC.....	35
3.5.2 Evaluasi SVM	36
BAB IV HASIL DAN PEMBAHASAN	38
4.1 Pengumpulan Data	38
4.2 <i>Text Preprocessing</i>	41
4.2.1 <i>Case Folding</i>	41
4.2.2 <i>Cleansing</i>	42
4.2.3 <i>Tokenization</i>	43
4.2.4 <i>Normalization</i>	43
4.2.5 <i>Stopword Removal</i>	44
4.2.6 <i>Stemming</i>	45
4.2.7 <i>Labelling</i>	45
4.2.8 TF-IDF	47
4.2.9 N-Gram	52
4.4 Klasifikasi	53
4.4.1 Klasifikasi <i>Naive Bayes Classifier</i>	54
4.4.2 Klasifikasi <i>Support Vector Machine</i>	57
4.5 Evaluasi.....	60
4.5.1 Evaluasi <i>Naive Bayes Classifier</i>	62
4.5.2 Evaluasi <i>Support Vector Machine</i>	64
4.6 Perbandingan NBC dan SVM	66
4.6.1 Perbandingan <i>Confusion Matrix</i>	66
4.6.2 Perbandingan Hasil Metrik.....	67
4.6.3 Perbandingan <i>K-Fold Cross Validation</i>	69
BAB V PENUTUP	71
5.1 Keterbatasan Penelitian.....	71
5.2 Kesimpulan	71
5.2.1 <i>Text Pre-Processing</i>	71
5.2.2 Sentimen Masyarakat.....	72

5.2.3 Perbandingan Kinerja NBC dan SVM	72
5.3 Saran.....	72
DAFTAR PUSTAKA.....	74

DAFTAR GAMBAR

Gambar 1.1 Peta Incidence Rate (IR) Kasus Dengue di Indonesia Tahun 2022	1
Gambar 2.1 Alur <i>Text Preprocessing</i>	11
Gambar 3.1 Alur Penelitian	23
Gambar 4.1 Upload Link Postingan Instagram di Apify	38
Gambar 4.2 Hasil Pengambilan Data di Apify	40
Gambar 4.3 Tampilan Data CSV	40
Gambar 4.4 Hasil <i>Labelling</i>	46
Gambar 4.5 <i>Wordcloud</i> TF-IDF	50
Gambar 4.6 Bar Chart TF-IDF	51
Gambar 4.7 Hasil SMOTE	54
Gambar 4.8 Hasil <i>Hyperplane</i>	60
Gambar 4.9 Evaluasi Performa NBC dan SVM	68
Gambar 4.10 Perbandingan Metrik Pada <i>K-Fold Cross Validation</i> NBC dan SVM	69

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	17
Tabel 2.2 Penelitian Sejenis	19
Tabel 4.1 Postingan Instagram	39
Tabel 4.2 Contoh Hasil Kumpulan Data	40
Tabel 4.3 Contoh Hasil <i>Casefolding</i>	42
Tabel 4.4 Contoh Hasil Cleansing.....	43
Tabel 4.5 Contoh Hasil <i>Tokenization</i>	43
Tabel 4.6 Contoh Hasil <i>Normalization</i>	44
Tabel 4.7 Contoh Hasil <i>Stopword Removal</i>	44
Tabel 4.8 Contoh Hasil <i>Stemming</i>	45
Tabel 4.9 Contoh Hasil <i>Labelling</i>	46
Tabel 4.10 Data Sampel dari Kolom <i>Stemming</i>	47
Tabel 4.11 Contoh Hasil TF-IDF	49
Tabel 4.12 Contoh Hasil N-Gram	52
Tabel 4.13 Kombinasi N-Gram dan TF-IDF.....	53
Tabel 4.14 Perbandingan Data Latih dan Data Uji.....	53
Tabel 4.15 Sampel Data Latih NBC.....	55
Tabel 4.16 Sampel Data Latih SVM	57
Tabel 4.17 Vektor Fitur SVM	57
Tabel 4.18 Perbandingan SMOTE pada <i>Confusion Matrix</i> NBC dan SVM.....	61
Tabel 4.19 <i>Confusion Matrix</i> NBC	62
Tabel 4.20 Hasil <i>K-Fold Cross Validation</i> NBC.....	63
Tabel 4.21 <i>Confusion Matrix</i> SVM.....	64
Tabel 4.22 Hasil <i>K-Fold Cross Validation</i> SVM.....	65
Tabel 4.23 Perbandingan <i>Confusion Matrix</i> NBC dan SVM.....	66
Tabel 4.24 Hasil Perhitungan Metrik dari <i>Confusion Matrix</i>	67
Tabel 4.25 Perbandingan <i>K-Fold Cross Validation</i> NBC dan SVM.....	69

DAFTAR LAMPIRAN

LAMPIRAN 1. Formulir Persetujuan Publikasi

LAMPIRAN 2. *Letter of Acceptance*

LAMPIRAN 3. Publikasi Ilmiah

LAMPIRAN 4. Bukti In Press

LAMPIRAN 5. Hasil Pengecekan Software Ithenticate/Turnitin

LAMPIRAN 6. Catatan Notulen Sidang Komprehensif

LAMPIRAN 7. Universitas Sriwijaya *English Proficiency Test*

DAFTAR ISTILAH

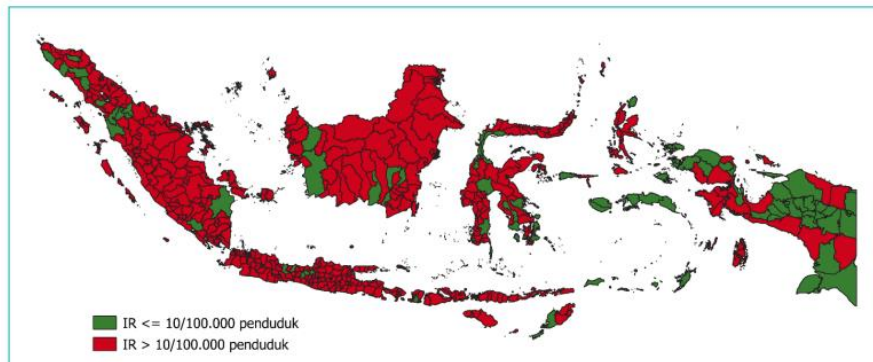
1. DBD : Demam Berdarah Dengue
2. NBC : Naive Bayes Classifier
3. SVM : Support Vector Machine
4. TF-IDF : Term Frequency-Inverse Document Frequency
5. SMOTE : Synthetic Minority Oversampling Technique
6. NLP : Natural Language Processing
7. VADER : Valence Aware Dictionary and Sentiment Reasoner
8. CSV : Comma-Separated Values
9. IR : Incidence Rate
10. TP : True Positive
11. TN : True Negative
12. FP : False Positive
13. FN : False Negative

BAB I

PENDAHULUAN

1.1 Latar Belakang

Demam Berdarah *Dengue* (DBD) merupakan masalah kesehatan serius di negara tropis seperti Indonesia, dengan angka kejadian yang terus meningkat setiap tahun. Pada Gambar 1.1 menunjukkan bahwa tahun 2022, terdapat 143 ribu kasus dengue di Indonesia (Kemenkes RI, 2022):



Gambar 1.1 Peta Incidence Rate (IR) Kasus Dengue di Indonesia Tahun 2022

Tujuan penanggulangan *dengue* adalah menurunkan beban kesehatan masyarakat dengan target 80% kabupaten/kota memiliki IR dengue $\leq 10/100.000$ penduduk pada tahun 2022 (Kemenkes RI, 2022). Namun, capaian IR *dengue* nasional pada tahun 2022 adalah 52/100.000, menunjukkan bahwa masih banyak pekerjaan yang harus dilakukan.

Untuk mengatasi penyebaran penyakit ini, salah satu strategi yang diimplementasikan adalah penggunaan nyamuk yang telah diinfeksi bakteri *Wolbachia* (Kemenkes RI, 2023). Program ini telah diimplementasikan di beberapa daerah di Indonesia dan mendapat perhatian luas dari masyarakat. Respons publik terhadap program ini dapat dilihat melalui media sosial, di mana masyarakat mengekspresikan pandangan, kekhawatiran, dan harapan mereka terkait efektivitas program tersebut.

Analisis sentimen merupakan alat yang penting untuk memahami opini publik di media sosial. Dengan analisis ini, kita dapat mengevaluasi sentimen positif atau negatif yang terkait dengan program *Wolbachia*. Untuk melakukan analisis sentimen ini, diperlukan metode klasifikasi yang andal. Dua metode klasifikasi yang sering digunakan adalah *Naive Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)* (Pang et al., 2002).

Berbagai penelitian telah membandingkan kinerja algoritma NBC dan SVM dalam konteks analisis sentimen, namun sebagian besar penelitian sebelumnya menggunakan Twitter sebagai platform analisis. Misalnya, penelitian Irwanto et al. (2023) menganalisis respons publik terhadap vaksinasi Covid-19 melalui Twitter, membandingkan NBC dengan *XGBoost Classifier*. Meskipun NBC menunjukkan kinerja yang baik dalam hal waktu proses dan lebih efisien dalam menangani data yang besar, namun proses pelabelan data yang manual berpotensi menyebabkan bias (Irwanto et al., 2023).

Bengesi et al. (2023) menganalisis sentimen publik terhadap wabah *monkeypox* menggunakan Twitter dan menemukan bahwa model dengan *TextBlob*, *Lemmatization*, *CountVectorizer*, dan SVM mencapai akurasi tertinggi sekitar 93,48%. SVM bekerja dengan baik pada data yang memiliki banyak fitur dan dapat bekerja dengan baik pada data yang belum pernah dilihat sebelumnya, begitu pula dengan NBC yang cepat dan efisien terutama pada data yang besar (Bengesi, et al., 2023). NBC juga mudah diimplementasikan serta tidak memerlukan banyak parameter untuk diatur. Namun, penelitian ini hanya mencakup data dari periode tertentu saja, yaitu Juli sampai dengan September tahun 2022 (Bengesi et al., 2023).

Penelitian dari Kusumawati et al. (2019) yang membandingkan kinerja NBC dan SVM dalam mengklasifikasikan opini pengguna Twitter tentang layanan Tokopedia, dengan hasil bahwa SVM lebih unggul. Pada penelitian tersebut, SVM lebih efektif menangani data yang kompleks, sementara NBC sangat sederhana karena hanya menggunakan frekuensi teks untuk menghitung probabilitas posteriornya. Selain itu, NBC memiliki tingkat akurasi dan kinerja yang tinggi dalam hal klasifikasi. Namun, penelitian ini memiliki keterbatasan pada jumlah data

yang terbatas dan hanya menggunakan dua kelas (positif dan negatif) (Kusumawati et al., 2019).

Penelitian oleh Reddy et al. (2024) membandingkan efektivitas Naive Bayes dan SVM dalam memprediksi tingkat kepuasan penumpang maskapai yang menggunakan ukuran sampel hanya 818 serta data dari dataset terbuka di Kaggle yang mungkin tidak mencerminkan kondisi nyata secara keseluruhan. Hasil menunjukkan bahwa SVM mencapai akurasi 94%, sementara Naive Bayes mencapai 86.8%, namun dengan sampel yang digunakan juga berisiko terhadap *overfitting* (Reddy et al., 2024).

Dari penelitian-penelitian yang dijelaskan sebelumnya, didapati beberapa kesamaan antara dua metode yang dipakai dalam penelitian ini, berikut merupakan kesamaan mendasar dari SVM dan NBC (Bengesi et al., 2023):

1. Kedua metode tersebut digunakan untuk melakukan klasifikasi, dimana keduanya mengkategorikan data ke dalam kelas yang telah ditentukan.
2. SVM dan NBC ini merupakan algoritma dari *supervised learning* yang artinya metode tersebut butuh data yang memiliki label untuk melatih model.
3. Kedua metode tersebut dapat digunakan dalam klasifikasi multikelas, walaupun SVM lebih sering dipakai dalam klasifikasi biner.
4. SVM dan NBC ini efektif dalam menangani data teks seperti dalam analisis sentimen atau klasifikasi dokumen.

Berdasarkan penelitian dari Mustakim et al. (2022) mengenai analisis sentimen menggunakan NBC dan SVM terhadap KAI *access*, kedua metode tersebut terbukti efektif dalam menangani data teks khususnya analisis sentimen berbasis aspek. Hasil dari penelitian tersebut menunjukkan bahwa SVM memberika performa terbaik dengan rata-rata *accuracy* 91.63%, *f1-score* 75.55%, *precision* 77.60% dan *recall* 74.47% (Mustakim et al., 2022). Sedangkan pada NBC lebih unggul pada *recall* terhadap data yang tidak seimbang, namun perforamnya secara keseluruhan lebih rendah dibandingkan dengan SVM (Mustakim et al., 2022).

Penelitian dari Laturiuw et al. (2023) menggunakan metode NBC dan SVM dalam analisis sentimen mengenai ulasan destinasi wisata Raja Ampat. Pada penelitian tersebut, SVM lebih unggul dalam menangani data teks berdimensi tinggi. Performa pada SVM juga lebih konsisten dalam memisahkan kelas yang positif dan negatif. Metode SVM memanfaatkan margin optimal untuk klasifikasi agar menghasilkan akurasi yang tinggi (Sari et al., 2023). Sedangkan NBC unggul dalam efisiensi proses klasifikasi dengan pendekatan probabilistik (Christanto et al., 2022). Meskipun NBC merupakan metode dengan kinerja yang sederhana, metode ini mampu memberikan hasil yang kompetitif, terutama pada data dengan struktur sederhana dan distribusi fitur yang jelas (Laturiuw et al., 2023).

Penelitian berikutnya yang menunjukkan bahwa NBC dan SVM efektif dalam menangani data teks yaitu penelitian dari Christanto et al. (2022) yang mengevaluasi kinerja dari NBC dan SVM dalam menganalisis sentimen ulasan pelanggan terhadap produk dan layanan hotel di Labuan Bajo. SVM menunjukkan kinerja unggul dalam memproses data teks yang kompleks, dengan kemampuan menangkap pola dari data berdimensi tinggi menggunakan hyperplane optimal (Nurmalasari et al., 2023). Metode ini konsisten menghasilkan akurasi yang tinggi, terutama pada data yang sulit dipisahkan secara linear (Christanto et al., 2022). Sementara itu, NBC menawarkan pendekatan yang sederhana namun efektif, dengan memanfaatkan probabilitas untuk mengklasifikasikan teks berdasarkan hubungan antar fitur (Christanto et al., 2022). Hasilnya, NBC memberikan waktu komputasi yang cepat dan akurasi yang layak, menjadikannya pilihan yang baik untuk dataset dengan struktur fitur yang sederhana (Christanto et al., 2022).

Penelitian-penelitian yang dipaparkan pada latar belakang ini memberikan dasar yang kuat untuk perbandingan kinerja NBC dan SVM. Namun pada sebagian besar penelitian tersebut menggunakan platform Twitter sebagai sumber data utama. Adapun platform sosial media lain seperti Instagram memiliki karakteristik interaksi visual dan naratif yang unik, belum banyak dieksplorasi. Instagram memungkinkan pengguna mengekspresikan sentimen mereka melalui teks dan visual, memberikan perspektif yang lebih mendalam dalam memahami reaksi publik terhadap program kesehatan seperti pelepasan nyamuk Wolbachia. Studi

terhadap virus Zika yang diteliti oleh Seltzer et al. (2017) sebelumnya menunjukkan bahwa platform ini dapat mengungkap ketakutan, misinformasi, dan dukungan publik yang signifikan (Seltzer et al., 2017).

Penelitian ini mengisi celah tersebut dengan fokus pada efektivitas kedua metode dalam konteks Instagram dan menggunakan data yang lebih luas dan representatif, dengan harapan dapat memberikan kontribusi signifikan dalam memahami dan memilih metode klasifikasi yang paling efektif untuk analisis sentimen terhadap penyebaran nyamuk *Wolbachia* di Indonesia.

1.2 Rumusan Masalah

Dari penjelasan di latar belakang maka terdapat beberapa rumusan masalah yang dibahas pada penelitian ini yaitu:

1. Bagaimana cara melakukan *preprocessing* data komentar netizen di Instagram tentang penyebaran nyamuk *Wolbachia* di Indonesia untuk mendapatkan data yang bersih dan siap digunakan untuk analisis sentimen?
2. Bagaimana hasil klasifikasi dari komentar netizen di Instagram tentang nyamuk *Wolbachia* di Indonesia dengan menggunakan algoritma NBC dan SVM?
3. Bagaimana perbandingan efektivitas dari NBC dan SVM dalam klasifikasi komentar netizen di Instagram tentang nyamuk *Wolbachia* di Indonesia?

1.3 Batasan Masalah

Batasan masalah dari penelitian yang dilakukan ini adalah dengan menggunakan kumpulan data komentar yang berbahasa Indonesia dan berformat teks melalui postingan @kemenkes_ri dan juga postingan dari akun lainnya di Instagram yang menyebarkan informasi mengenai nyamuk *Wolbachia* ini.

1.4 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan ini yaitu:

1. Melakukan *preprocessing* data komentar netizen di Instagram tentang penyebaran nyamuk *Wolbachia* di Indonesia untuk analisis sentimen.
2. Menganalisis hasil klasifikasi dari komentar netizen di Instagram tentang penyebaran nyamuk *Wolbachia* di Indonesia dengan menggunakan algoritma

NBC dan SVM.

3. Membandingkan efektivitas dari NBC dan SVM dalam klasifikasi komentar netizen di Instagram tentang penyebaran nyamuk *Wolbachia* di Indonesia.

1.5 Manfaat Penelitian

Analisis sentimen nyamuk *Wolbachia* ini memberikan manfaat serta kontribusi bagi berbagai pihak seperti berikut:

1. Hasil penelitian ini dapat membantu peneliti serta pemerintah untuk meningkatkan kualitas dan efektivitas dari program penyebaran nyamuk *Wolbachia* ini untuk mengurangi penyakit DBD.
2. Data sentimen dari penelitian ini dapat digunakan untuk menginformasikan kebijakan dan strategi komunikasi yang lebih baik.
3. Memberikan gambaran yang jelas kepada masyarakat mengenai manfaat dan tujuan dari program penyebaran nyamuk *Wolbachia* ini untuk mengurangi miskonsepsi dan kekhawatiran yang ada.
4. Penelitian ini dapat berkontribusi pada pengembangan metodologi dalam analisis sentimen, khususnya dalam penggunaan algoritma NBC dan SVM.
5. Menambah literatur ilmiah tentang analisis sentimen di media sosial, khususnya mengenai kesehatan masyarakat.
6. Memberikan wawasan lebih dalam tentang bagaimana masyarakat bereaksi terhadap inovasi teknologi dalam kesehatan, yang dapat digunakan untuk penelitian lebih lanjut di bidang serupa.

DAFTAR PUSTAKA

- Aathira, S., Reena, B., & Uma, D. S. (2023). Mental Stress Detection Using TF-IDF with Multinomial Naive Bayes. *International Journal for Research in Science Engineering & Technology (IJRSET)*, 9(6), 1–5.
- Astuti, F. D., & Lenti, F. N. (2021). Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution Menggunakan KNN. *JUPITER*, 13, 89–98.
- Bengesi, S., Oladunni, T., Olusegun, R., & Audu, H. (2023). A Machine Learning-Sentiment Analysis on Monkeypox Outbreak : An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets. *IEEE Access*, 11(January), 11811–11826. <https://doi.org/10.1109/ACCESS.2023.3242290>
- Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, 29, 509–553. <https://doi.org/10.1017/S1351324922000213>
- Christanto, H. J., & Singgalen, Y. A. (2022). Sentiment Analysis of Customer Feedback Reviews Towards Hotel's Products and Services in Labuan Bajo. *Journal of Information Systems and Informatics*, 4(4), 805–822. <https://doi.org/10.51519/journalisi.v4i4.294>
- Dwi Putra, H., Khairani, L., Hastari, D., Studi Sistem Informasi, P., Sains dan Teknologi, F., & Author, C. (2023). Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 120–125. <https://journal.irpi.or.id/index.php/sentimas>
- Hadi, K., & Utami, E. (2024). *Analysis of K-NN with the Integration of Bag of Words , TF-IDF , and N-Grams for Hate Speech Classification on Twitter*. 12(2), 289–298.
- Hafid, H. (2023). Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia. *Journal of Mathematics*, 6(2), 161–168. <http://www.ojs.unm.ac.id/jmathcos>

- Hilda Kusumahadi, S., Junaedi, H., & Santoso, J. (2019). Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 54–60. <https://doi.org/10.30591/jpit.v4i1.1125>
- Hoffmann, A. (2020). Quick guide Wolbachia Cryptomonads. *Current Biology*. <https://doi.org/10.1016/j.cub.2020.08.039>
- Hormansyah, D. S., & Aulia, I. (2018). Penerapan Metode TF-IDF dan N-Gram pada Pengembangan Aplikasi Chatbot Berbasis Line Untuk Layanan Publik Kesehatan di Kota Malang. *Jurnal Informatika Polinema*, 5, 7–11.
- Hu, L. F., Gong, W., Qi, L. X., & Wang, P. (2013). *A Method for Feature Selection Based On the Optimal Hyperplane of SVM and Independent Analysis*. 14–17. <https://doi.org/10.1109/ICMLC.2013.6890454>
- Hugo, L. E., Rasic, G., Maynard, A. J., Ambrose, L., Liddington, C., Thomas, C. J. E., Nath, N. S., Graham, M., Winterford, C., Wimal Siri-Yapa, B. M. C. R., Xi, Z., Beebe, N. W., & Devine, G. J. (2022). *Wolbachia wAlbB inhibit dengue and Zika infection in the mosquito Aedes aegypti with an Australian background*. <https://doi.org/https://doi.org/10.1101/2022.03.22.485408>
- Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/teknika.v10i1.311>
- Irwanto, A., & Goeirmento, L. (2023). *Sentiment Analysis from Twitter about Covid-19 Vaccination in Indonesia using Naïve Bayes and XGboost Classifier Algorithm*. 27(2), 145–152.
- Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques. *IEEE Access*, 9, 39707–39716. <https://doi.org/10.1109/ACCESS.2021.3064084>
- Jamshed, H., Khan, M. S. A., Khurram, M., Inayatullah, S., & Athar, S. (2019). Data

- Preprocessing: A Preliminary Step for Web Data Mining. *3C Tecnologia, May*, 206–221. [https://doi.org/http://dx.doi.org/10.17993/3ctecno.2019.specialissue2.206–221](https://doi.org/http://dx.doi.org/10.17993/3ctecno.2019.specialissue2.206-221) Suggested
- Kemenkes RI. (2022). Membuka Lembaran Baru. *Laporan Tahunan 2022 Demam Berdarah Dengue*, 17–19.
- Kemenkes RI. (2023). Inovasi Wolbachia, Cara Ampuh dan Hemat Kendalikan Demam Berdarah. *Kementerian Kesehatan RI*, 1–2. <https://www.kemkes.go.id/id/rilis-kesehatan/inovasi-wolbachia-cara-ampuh-dan-hemat-kendalikan-demam-berdarah>
- Khotimah, A. C., & Utami, E. (2022). Comparison Naive Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine in the classification of individual on twitter account. *Jurnal Teknik Informatika (JUTIF)*, 3(3), 673–680. <https://doi.org/10.20884/1.jutif.2022.3.3.254>
- Kusumawati, R., Arofah, D, A., & Pramana, P, A. (2019). Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter ' s Classification of Tokopedia Services. *Journal of Physics: Conference Series*, 1320, 1–11. <https://doi.org/10.1088/1742-6596/1320/1/012016>
- Laturiuw, A. K., & Singgalen, Y. A. (2023). Sentiment Analysis of Raja Ampat Tourism Destination Using CRISP-DM: SVM, NBC, DT, and k-NN Algorithm. *Journal of Information Systems and Informatics*, 5(2), 518–535. <https://doi.org/10.51519/journalisi.v5i2.490>
- Mardjo, A., & Choksuchat, C. (2022). HyVADRF: Hybrid VADER-Random Forest and GWO for Bitcoin Tweet Sentiment Analysis. *IEEE Access*, 10(August), 101889–101897. <https://doi.org/10.1109/ACCESS.2022.3209662>
- Maulana, R., Rahayuningsih, P. A., Irmayani, W., Saputra, D., & Jayanti, W. E. (2020). Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012060>

- Muhaddisi, A., Prastowo, B. N., & Putri, D. U. K. (2021). Sentiment Analysis With Sarcasm Detection On Politician's Instagram. *Sentiment Analysis With Sarcasm Detection On Politician's Instagram*, Vol.15, No(October 2021), 349–358. <https://doi.org/https://doi.org/10.22146/ijccs.66375>
- Mustakim, H., & Priyanta, S. (2022). Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 16(2), 113. <https://doi.org/10.22146/ijccs.68903>
- Narayan, S., & Sathiyamoorthy, E. (2022). Early Prediction of Heart Diseases using Naive Bayes Classification Algorithm and Laplace Smoothing Technique. *International Journal of Grid and High Performance Computing*, 14(1), 1–14. <https://doi.org/10.4018/IJGHPC.316157>
- Navin Kumar R, & Sneha, S. (2022). Multi-Strategy Sentiment Analysis of Consumer Reviews with Partial Phrase Matching. *International Journal of Computer Techniques*, 9(2).
- Nurmalasari, D., Hermanto, T. I., & Nugroho, I. M. (2023). Perbandingan Algoritma SVM, KNN dan NBC Terhadap Analisis Sentimen Aplikasi Loan Service. *Jurnal Media Informa*, 7, 1521–1530. <https://doi.org/10.30865/mib.v7i3.6427>
- Osterholz, S., Mosel, E. I., & Egloff, B. (2022). #Insta personality: Personality expression in Instagram accounts, impression formation, and accuracy of personality judgments at zero acquaintance. *Journal of Personality*, 91, 566–582.
- Petiwi, M. I., Triayudi, A., & Sholihati, I. D. (2022). Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, 6(1), 542. <https://doi.org/10.30865/mib.v6i1.3530>
- Pramana, D., Afdal, M., Mustakim, M., & Permana, I. (2023). Analisis Sentimen Terhadap Pemindahan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbors. *Jurnal Media Informatika Budidarma*, 7(3), 1306–1314. <https://doi.org/10.30865/mib.v7i3.6523>

- Putra, D., & Wibowo, A. (2020). Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2, 84–92.
- Putra Negara, A. B. (2023). The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, 14(3), 172. <https://doi.org/10.24843/lkjiti.2023.v14.i03.p05>
- Reddy, E. P. V., & Ramesh, S. (2024). Comparative analysis of support vector machine and Naive Bayes classifier for the prediction of airline passenger satisfaction. *AIP Conference Proceedings*.
- Riadi, S., Utami, E., & Yaqin, A. (2023). Comparison of NB and SVM in Sentiment Analysis of Cyberbullying using Feature Selection. *Sinkron*, 8(4), 2414–2424. <https://doi.org/10.33395/sinkron.v8i4.12629>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5, 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Sari, A. W., Hermanto, T. I., & Defriani, M. (2023). Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine. *Sinkron*, 8(3), 1366–1378. <https://doi.org/10.33395/sinkron.v8i3.12447>
- Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1–47. <https://doi.org/10.1145/505282.505283>
- Seltzer, E. K., Lu, M., & Merchant, R. M. (2017). *Public sentiment and discourse about Zika virus on Instagram*. 0(215), 0–5.
- Setiawan, H., Utami, E., & Sudarmawan, S. (2021). Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, 5(1), 43–51. <https://doi.org/10.31603/komtika.v5i1.5189>
- Singgih, M. N. A., & Utari, D. T. (2024). The amalgamation of XGBoost and SMOTE for sentiment analysis pedulilindung application reviews. *AIP Conference*

Proceedings, 3027(1). <https://doi.org/10.1063/5.0204739>

- Suryani, S., Fayyad, M. F., Savra, D. T., Kurniawan, V., & Estanto, B. H. (2023). Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm. *Public Research Journal of Engineering, Data Technology and Computer Science*, 1(1), 1–9. <https://doi.org/10.57152/predatecs.v1i1.814>
- Tembusai, Z. R., Mawengkang, H., & Zarlis, M. (2021). K-Nearest Neighbor with K-Fold Cross Validation and Analytic Hierarchy Process on Data Classification. *International Journal of Advances in Data and Information Systems*, 2(1), 1–8. <https://doi.org/10.25008/ijadis.v2i1.1204>
- Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, 57, 117–126. <https://doi.org/10.1016/j.eswa.2016.03.028>
- Xu, S. (2018). Bayesian Naive Bayes Classifiers to Text Classification. *Journal of Information Science*, 44(15), 48–50. <https://doi.org/10.1177/0165551516677946>
- Zain, H. H., Awangga, R. M., & Rahayu, W. I. (2023). Perbandingan Model Svm, Knn Dan Naïve Bayes Untuk Analisis Sentiment Pada Data Twitter: Studi Kasus Calon Presiden 2024. *JIMPS: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, 8(3), 2083–2093. <https://jim.usk.ac.id/sejarah>
- Židek, K., Hošovský, A., & Dubjak, J. (2016). Diagnostics of surface errors by embedded vision system and its classification by machine learning algorithms. *Key Engineering Materials*, 669, 459–466. <https://doi.org/10.4028/www.scientific.net/KEM.669.459>