

PEMETAAN PERSEPSI PUBLIK TERHADAP TRANSFORMASI DIGITAL PERTANIAN MENGUNAKAN MULTINOMIAL NAIVE BAYES DAN TF-IDF

Betran Alfredo Simanjuntak¹, Hotler Manurung², Kristina Annatasia Br Sitepu³

Sistem Informasi, STMIK Kaputama Binjai, Sumatra Utara

E-mail: [*bertran2611@gmail.com](mailto:bertran2611@gmail.com)¹, manurunghotler0@gmail.com², kannatasia88@gmail.com³

ABSTRAK

Sektor pertanian Indonesia tengah bertransformasi menuju era digitalisasi guna meningkatkan efisiensi dan ketahanan pangan nasional, namun implementasinya menghadapi tantangan berupa kesenjangan kesiapan serta keberagaman persepsi masyarakat. Permasalahan utama dalam penelitian ini terletak pada kompleksitas ekstraksi informasi strategis dari ribuan opini publik yang masif dan tidak terstruktur di media sosial. Penelitian ini bertujuan untuk memetakan persepsi publik terhadap transformasi digital pertanian menggunakan metode *Multinomial Naive Bayes* yang diintegrasikan dengan teknik pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF). Data penelitian diperoleh melalui teknik *scraping* terhadap 3.268 komentar di platform YouTube, yang selanjutnya diolah menggunakan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dan pengujian *relevancy filtering*. Hasil pengujian menunjukkan bahwa tahap prapemrosesan mampu mereduksi *noise* sebesar 64,6%, menghasilkan 1.156 data valid untuk fase pemodelan. Model klasifikasi yang dibangun berhasil mencapai tingkat akurasi sebesar 83,39%, dengan rincian distribusi sentimen didominasi oleh opini positif sebanyak 292 data, sentimen negatif 237 data, dan sentimen netral 113 data. Kesimpulan utama penelitian ini menegaskan bahwa kombinasi *Multinomial Naive Bayes* dan TF-IDF memiliki performa yang andal sebagai instrumen pemetaan opini digital guna mendukung evaluasi kebijakan serta perumusan strategi pemangku kepentingan dalam modernisasi agrikultur Indonesia secara objektif.

Kata kunci

Analisis Sentimen, Digitalisasi Pertanian, Komentar YouTube, *Multinomial Naive Bayes*, Pemetaan Persepsi.

ABSTRACT

Indonesia's agricultural sector is undergoing a digital transformation to enhance efficiency and national food security; however, its implementation faces challenges stemming from technological readiness gaps and diverse public perceptions. The core issue addressed in this study is the complexity of extracting strategic insights from massive, unstructured social media data for accurate policy evaluation. This research aims to map public perceptions of agricultural digital transformation using the *Multinomial Naive Bayes* algorithm integrated with *Term Frequency-Inverse Document Frequency* (TF-IDF) feature weighting. Grounded in the CRISP-DM framework, 3,268 comments were scraped from YouTube and subjected to rigorous preprocessing and relevancy filtering. The results reveal that relevancy filtering successfully reduced noise by 64.6%, yielding 1,156 valid data points for modeling. The resulting classification model achieved an overall accuracy of 83.39%, with sentiment distribution dominated by positive opinions (292 entries), followed by negative (237 entries) and neutral (113 entries) sentiments. This study concludes that combining *Multinomial Naive Bayes* and TF-IDF provides a reliable digital opinion-mapping instrument, supporting objective policy evaluation and stakeholder strategy formulation for modernizing Indonesian agriculture.

Keywords

Sentiment Analysis, Agricultural Digitalization, YouTube Comments, *Multinomial Naive Bayes*, Perception Mapping.

1. PENDAHULUAN

Sektor pertanian merupakan pilar strategis pembangunan nasional Indonesia yang memiliki peran vital sebagai penyedia pangan dan penopang stabilitas ekonomi. Saat ini, transformasi menuju era *Digital Agriculture* tengah didorong secara masif melalui integrasi teknologi seperti *Internet of Things* (IoT), aplikasi manajemen tani, hingga platform pemasaran digital untuk meningkatkan efisiensi dari hulu ke hilir. Namun, keberhasilan implementasi teknologi ini sangat bergantung pada tingkat penerimaan dan kesiapan masyarakat petani sebagai pengguna utama. Di tengah arus digitalisasi, media sosial seperti YouTube telah bertransformasi menjadi ruang publik digital di mana masyarakat secara spontan menyuarakan aspirasi, keluhan, maupun dukungan terhadap kebijakan pemerintah. Sejalan dengan hal tersebut, Prabowo et al. (2026) menegaskan bahwa platform digital ini menyimpan kekayaan informasi mengenai suasana batin masyarakat (*public mood*), yang jika dikelola dengan teknik *Natural Language Processing* (NLP), dapat diubah menjadi wawasan berharga untuk mengevaluasi dampak kebijakan pemerintah secara objektif. Fenomena ini menciptakan pangkalan data opini yang besar, di mana setiap komentar mencerminkan persepsi jujur publik terhadap realitas transformasi teknologi di lapangan.

Meskipun potensi data media sosial sangat besar, tantangan utama yang dihadapi oleh pemangku kebijakan adalah sulitnya mengekstraksi informasi strategis dari ribuan opini yang tidak terstruktur dan berjumlah masif. Tanpa adanya instrumen pemetaan yang sistematis, evaluasi terhadap efektivitas kebijakan digitalisasi pertanian berisiko tidak tepat sasaran. Masalah ini diperumit oleh rendahnya literasi digital dan keterbatasan infrastruktur di wilayah perdesaan yang seringkali memicu sentimen negatif atau resistensi. Pihak yang paling terdampak adalah pemerintah dan pengembang teknologi yang kehilangan basis data objektif untuk menentukan apakah suatu program perlu dilanjutkan, diperbaiki, atau diubah strateginya sesuai kebutuhan nyata petani. Oleh karena itu, penelitian mengenai analisis sentimen menjadi sangat krusial sebagai jembatan antara kebijakan teknologi dan realitas persepsi masyarakat.

Penelitian terdahulu telah berupaya membedah opini publik di media sosial dengan berbagai pendekatan. Hidayat, Bahtiar, dan Kaslani (2025) berhasil mengimplementasikan algoritma Naive Bayes untuk menganalisis sentimen teknologi 5G pada YouTube dengan tingkat akurasi tinggi. Selaras dengan hal tersebut, Ardiansyah et al. (2025) memvalidasi bahwa Naive Bayes memiliki efisiensi waktu komputasi yang lebih unggul dibandingkan *Random Forest* dalam memproses data teks media sosial yang masif. Dalam domain pertanian, Umaruddin, Febriansyah, dan Tahira (2025) melakukan analisis konten terhadap representasi isu agrikultur di YouTube dan menemukan dinamika respons audiens yang sangat bergantung pada sub-tema yang dibahas. Sementara itu, Ma'rufudin dan Yudhistira (2025) menggunakan *Support Vector Machine* (SVM) untuk memotret sentimen petani milenial di platform X, membuktikan bahwa jejak digital petani saat ini sudah dapat diukur secara sistematis.

Meskipun penelitian-penelitian di atas memberikan landasan metodologis yang kuat, terdapat beberapa keterbatasan signifikan yang perlu dikaji secara kritis. Sebagian besar studi analisis sentimen sebelumnya masih terfokus pada isu teknologi umum, layanan komersial, atau kebijakan sosial secara makro, namun sangat terbatas yang secara spesifik membedah fenomena digitalisasi pertanian di Indonesia. Selain itu, karakteristik bahasa informal di YouTube yang penuh dengan *slang*, singkatan, dan *noise* seringkali menjadi kendala yang menurunkan performa model jika tidak ditangani

dengan prapemrosesan yang intensif. Banyak penelitian terdahulu mengabaikan tahap pembersihan relevansi yang ketat, sehingga hasil klasifikasi berisiko terdistorsi oleh pesan iklan atau *spam* yang mendominasi kolom komentar media sosial. Kesenjangan ini menunjukkan perlunya sebuah model yang tidak hanya akurat secara algoritma, tetapi juga ketat dalam penyaringan kualitas data awal.

Research gap utama dalam penelitian ini terletak pada minimnya literasi ilmiah yang memetakan persepsi publik terhadap transformasi digital pertanian Indonesia dengan kerangka kerja data mining yang komprehensif. Kebaruan (*novelty*) penelitian ini ditunjukkan melalui penerapan teknik *relevancy filtering* yang sangat ketat berbasis kamus kata kunci khusus domain pertanian, yang terbukti mampu mereduksi *noise* hingga sebesar 64,6%. Selain itu, penelitian ini mengintegrasikan algoritma Multinomial Naive Bayes dengan teknik pembobotan TF-IDF untuk menangani karakteristik bahasa informal media sosial secara lebih presisi. Berbeda dengan penelitian sebelumnya yang cenderung generalis, studi ini menggunakan kerangka kerja CRISP-DM yang sistematis mulai dari tahap *business understanding* hingga *deployment* aplikasi web berbasis Streamlit sebagai instrumen monitoring opini publik.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membangun model klasifikasi otomatis yang mampu memetakan distribusi persepsi masyarakat ke dalam kategori positif, negatif, dan netral terkait transformasi digital agrikultur. Kontribusi akademik dari penelitian ini adalah memperkaya literatur mengenai optimalisasi Naive Bayes dalam pengolahan teks bahasa Indonesia informal pada domain spesifik pertanian. Secara praktis, hasil penelitian ini memberikan signifikansi bagi pemangku kebijakan sebagai bahan evaluasi objektif untuk memahami hambatan adopsi teknologi di level akar rumput. Melalui pemetaan sentimen ini, diharapkan pemerintah dapat merumuskan strategi digitalisasi yang lebih akomodatif dan sesuai dengan ekspektasi serta kesiapan masyarakat petani di Indonesia.

2. METODE PENELITIAN

2.1 Desain Penelitian

Penelitian ini merupakan jenis penelitian kuantitatif dengan pendekatan eksperimental laboratorium komputer yang berfokus pada bidang *Data Mining* dan *Natural Language Processing* (NLP). Desain penelitian mengadopsi kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang mencakup enam tahapan sekuensial: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Kerangka CRISP-DM dipilih karena menyediakan metodologi terstruktur dan terukur untuk mengekstraksi pengetahuan ilmiah dari pangkalan data teks tidak terstruktur.

2.2 Objek dan Ruang Lingkup

Objek penelitian ini adalah persepsi publik Indonesia terhadap fenomena transformasi digital di sektor agrikultur. Unit analisis data berupa komentar individu yang diunggah pengguna pada platform YouTube. Ruang lingkup penelitian dibatasi pada komentar berbahasa Indonesia yang secara spesifik membahas isu teknologi pertanian seperti *smart farming*, penggunaan *drone*, dan aplikasi manajemen tani. Penelitian ini berfokus pada klasifikasi polaritas sentimen dan tidak melakukan deteksi emosi mendalam maupun analisis sarkasme.

2.3 Data dan Sumber Data

Sumber data primer diperoleh dari kolom komentar kanal YouTube relevan seperti YESS Programme, CapCapung, dan saluran resmi kementerian terkait. Jenis data yang diolah adalah data teks tidak terstruktur. Data mentah yang dikumpulkan berjumlah 3.268 komentar, yang selanjutnya disaring melalui tahapan *relevancy filtering* hingga menghasilkan 1.156 data valid untuk pemodelan. Data mencakup rentang interaksi pengguna dari Juni 2025 hingga Januari

2026. Kriteria inklusi data valid mewajibkan keberadaan kata kunci spesifik domain pertanian (seperti "pupuk", "lahan", "teknologi") serta bebas dari pesan promosi maupun spam.

2.4 Variabel Penelitian

Variabel utama dalam penelitian ini terdiri atas Sentimen Publik sebagai variabel kategorikal dan Bobot TF-IDF sebagai variabel fitur numerik.

Tabel 1. Definisi Operasional Variabel Penelitian

Variabel	Definisi Operasional	Indikator	Skala	Sumber Data
Sentimen Publik	Kecenderungan opini masyarakat terhadap digitalisasi pertanian	Positif, Negatif, Netral	Nominal	YouTube API
Bobot TF-IDF	Signifikansi bobot kata dalam dokumen relatif terhadap korpus	Skor numerik TF x IDF	Rasio	Hasil Ekstraksi

2.5 Teknik Pengumpulan Data

Data dikumpulkan melalui teknik *scraping* otomatis memanfaatkan protokol YouTube Data API v3 berbasis bahasa pemrograman Python. Parameter ekstraksi data mencakup identitas unik video (*video_id*) dan batas pengambilan data (*max_results*). Data atribut yang berhasil diekstraksi meliputi nama pengguna, isi komentar, jumlah suka (*like count*), dan stempel waktu (*timestamp*) yang disimpan dalam format *Comma-Separated Values* (CSV).

2.6 Tahapan Penelitian

Alur pelaksanaan penelitian dirancang sistematis melalui tahapan berikut:

1. *Data Collection*: Penarikan data komentar publik via YouTube Data API v3.
2. *Relevancy Filtering*: Penyaringan berbasis kamus (*dictionary-based filtering*) untuk mereduksi *noise* berupa spam dan promosi.
3. *Text Preprocessing*: Pembersihan teks melalui tahap *case folding*, *cleaning*, *tokenizing*, *normalization* (penanganan bahasa gaul/slang), *stopword removal*, serta *stemming* menggunakan pustaka Sastrawi.
4. *Feature Engineering*: Transformasi data teks menjadi representasi vektor numerik menggunakan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF).
5. *Data Splitting*: Pembagian dataset dengan proporsi 70% sebagai data latih (*training set*) dan 30% sebagai data uji (*test set*).
6. *Training & Modeling*: Pelatihan model memanfaatkan algoritma *Multinomial Naive Bayes* (MNB).
7. *Evaluation*: Pengujian performa model berbasis *Confusion Matrix*.
8. *Deployment*: Perancangan dasbor interaktif sebagai sarana visualisasi persepsi publik.

2.7 Pemodelan Multinomial Naive Bayes

Metode klasifikasi yang diimplementasikan adalah *Multinomial Naive Bayes* yang dikombinasikan dengan fitur TF-IDF. Formulasi probabilitas posterior dihitung berdasarkan teorema Bayes :

$$P(C_k|X) = \frac{P(X|C_k) \cdot P(C_k)}{P(X)} \quad (2.1)$$

Untuk menghindari masalah probabilitas nol (*zero probability problem*), estimasi *likelihood* menggunakan *Laplace Smoothing* (+1) dihitung melalui persamaan:

$$P(w_i|C_k) = \frac{\text{count}(w_i, C_k) + 1}{\sum_{(w \in V)} \text{count}(w, C_k) + |V|} \quad (2.2)$$

Di mana w_i adalah kata ke- i , C_k adalah kelas sentimen, $\text{count}(w_i, C_k)$ adalah frekuensi kemunculan kata w_i pada kelas C_k , dan $|V|$ merupakan jumlah kata unik dalam seluruh kosakata korpus.

2.8 Teknik Evaluasi

Keandalan model diukur menggunakan matriks evaluasi *Confusion Matrix*. Metrik pengujian mencakup :

1. Akurasi (*Accuracy*): Rasio kebenaran prediksi total terhadap seluruh data uji.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.3)$$

2. Presisi (*Precision*): Tingkat ketepatan antara data yang diprediksi positif dengan data aktual.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2.4)$$

3. Daya Ingat (*Recall*): Kemampuan model dalam menemukan kembali informasi kelas positif secara keseluruhan.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2.5)$$

4. F1-Score: Rata-rata harmonik menyeimbang antara nilai presisi dan *recall*.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.6)$$

2.9 Lingkungan Perangkat Penelitian

Pengembangan eksperimen dilakukan menggunakan spesifikasi sistem sebagai berikut :

1. Bahasa Pemrograman: Python v3.10+
2. Lingkungan Pengembangan (IDE): Visual Studio Code
3. Pustaka Utama: Pandas & NumPy (manipulasi data), Scikit-learn (ekstraksi TF-IDF, MNB, dan metrik evaluasi), Sastrawi (stemmer Bahasa Indonesia), NLTK (proses tokenisasi & stopword).
4. Platform Deployment: Streamlit untuk membangun antarmuka dasbor pemantauan sentimen berbasis web.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan dan Filtrasi Data

Data primer dalam penelitian ini diperoleh melalui teknik *web scraping* otomatis terhadap kolom komentar pada kanal-kanal YouTube bertema agrikultur digital seperti YESS Programme, CapCapung, dan kementerian terkait selama periode Juni 2025 hingga Januari 2026. Proses akuisisi data awal menghasilkan total 3.268 komentar mentah. Mengingat karakteristik media sosial yang terbuka, data tersebut masih mengandung banyak *noise* berupa pesan iklan, promosi komersial, dan komentar di luar konteks.

Untuk menjaga validitas penelitian, diterapkan *relevancy filtering* berbasis kamus kata kunci (*dictionary-based*) khusus domain pertanian. Hasil penyaringan menunjukkan bahwa sebanyak 2.112 data (64,63%) dinyatakan tidak valid, sementara 1.156 data (35,37%) dinyatakan sebagai data valid yang merepresentasikan opini riil publik terhadap digitalisasi pertanian. Tingginya persentase data tidak valid menegaskan urgensi tahap filtrasi guna menghindari distorsi pola bahasa pada pemodelan klasifikasi.

Tabel 1. Perincian Hasil Pengumpulan dan Filtrasi Data

Kategori Data	Jumlah (Komentar)	Persentase (%)
Data Valid (Relevan)	1.156	35,37%
Data Tidak Valid (<i>Noise</i>)	2.112	64,63%
Total Dataset Mentah	3.268	100,00%

3.2 Hasil Prapemrosesan Data (*Text Preprocessing*)

Tahap prapemrosesan teks dilakukan secara sekuensial untuk mentransformasi data teks informal YouTube menjadi format yang konsisten. Tahapan ini meliputi *case folding*, *cleaning*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming* menggunakan pustaka Sastrawi. Proses normalisasi berperan krusial dalam mengubah kosakata tidak baku dan *slang* media sosial menjadi bentuk standar agar algoritma dapat mengekstraksi fitur secara presisi.

Tabel 2. Sampel Transformasi Data pada Tahap *Preprocessing*

Tahapan	Teks Sebelum	Teks Sesudah
Cleaning	"...dron masih belum terjangkau hrgnya bagi petani."	"...dron masih belum terjangkau hrgnya bagi petani"
Normalization	"...hrngnya bagi petani"	"...harganya bagi petani"

Stemming	"...harganya bagi petani"	"...harga bagi tani"
----------	---------------------------	----------------------

3.3 Hasil Ekstraksi Fitur (*Feature Engineering*)

Setelah data bersih, dilakukan ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Teknik ini memberikan bobot numerik pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen relatif terhadap seluruh korpus. Hasil analisis menunjukkan bahwa kata-kata spesifik yang membawa muatan sentimen seperti "efisien", "mahal", atau "canggih" mendapatkan bobot IDF yang lebih tinggi dibandingkan kata-kata umum seperti "pertanian" yang muncul di hampir seluruh dokumen. Transformasi vektor ini memungkinkan algoritma *Naive Bayes* melakukan perhitungan probabilitas posterior secara optimal.

3.4 Hasil Pemodelan

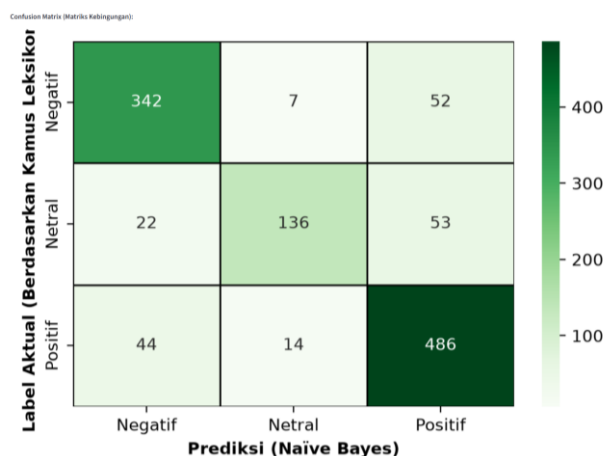
Pemodelan dilakukan dengan mengimplementasikan algoritma *Multinomial Naive Bayes* (MNB) dengan parameter *Laplace smoothing* (α) sebesar 0,1 untuk meminimalkan dampak probabilitas nol pada kata yang jarang muncul. Dataset valid (1.156 data) dipecah menjadi data latih (70%) untuk membangun basis pengetahuan probabilitas dan data uji (30%) untuk mengevaluasi kemampuan generalisasi model. Integrasi antara bobot statistik TF-IDF dan representasi leksikon dalam pemodelan ini dirancang untuk menangkap kompleksitas makna dalam opini publik.

3.5 Hasil Evaluasi Performa Model

Keandalan model divalidasi menggunakan instrumen *Confusion Matrix*. Berdasarkan pengujian empiris, model berhasil mencapai tingkat Akurasi Keseluruhan sebesar 83,39%. Performa ini melampaui ambang batas keberhasilan dasar (75%), menunjukkan bahwa model sangat akomodatif terhadap karakteristik bahasa informal media sosial.

Tabel 3. Metrik Evaluasi Klasifikasi Model *Multinomial Naive Bayes*

Kelas Sentimen	Presisi (Precision)	Sensitivitas (Recall)	F1-Score
Positif	0,822	0,893	0,856
Negatif	0,838	0,853	0,845
Netral	0,866	0,645	0,739
Rata-rata Akurasi	0,834	0,834	0,834



Gambar 1. Visualisasi Matriks Kebingungan Hasil Prediksi Sentimen Publik

3.6 Pembahasan Akademis dan Implikasi

Capaian akurasi sebesar 83,39% mencerminkan bahwa penerapan *relevancy filtering* yang ketat berhasil meminimalisir distorsi data, sehingga model mampu mempelajari pola sentimen secara objektif. Distribusi sentimen menunjukkan dominasi sentimen positif (292 data), yang merefleksikan optimisme masyarakat terhadap manfaat efisiensi yang ditawarkan oleh ekosistem *smart farming* dan penggunaan *drone*. Hal ini menjawab permasalahan penelitian mengenai tren penerimaan teknologi digital di tingkat akar rumput.

Namun, tingginya angka sentimen negatif (237 data) memberikan indikasi adanya hambatan struktural yang signifikan. Berdasarkan analisis kata kunci pada *Word Cloud*, keluhan publik berpusat pada tingginya modal awal pengadaan teknologi serta keterbatasan infrastruktur internet di wilayah perdesaan. Implikasinya, meskipun publik terbuka terhadap inovasi, pemerintah perlu melakukan intervensi kebijakan berupa subsidi teknologi atau pemerataan jaringan digital guna mengurangi resistensi. Model klasifikasi ini telah membuktikan perannya sebagai instrumen *monitoring* opini publik yang efektif untuk mendukung evaluasi kebijakan secara sistematis.

3.7 Kelebihan dan Keunikan Penelitian

Penelitian ini memiliki beberapa poin keunggulan dan kebaruan (*novelty*) yang kuat:

- a. Keunikan Objek dan Dataset: Mengkaji fenomena digitalisasi pertanian di Indonesia menggunakan dataset interaksi nyata di YouTube, di saat penelitian serupa lebih banyak berfokus pada isu teknologi umum atau media sosial Twitter/X.
- b. Kombinasi Metode: Penggunaan kerangka kerja CRISP-DM yang sistematis dengan integrasi pembobotan TF-IDF dan *Multinomial Naive Bayes* menghasilkan performa akurasi tinggi (83,39%) pada korpus bahasa informal.
- c. Inovasi Teknik Filtrasi: Implementasi *relevancy filtering* berbasis kamus yang mampu mereduksi *noise* hingga 64,63% merupakan kontribusi teknis yang menjamin validitas hasil klasifikasi.
- d. Kontribusi Praktis: Hasil penelitian diimplementasikan ke dalam antarmuka web berbasis Streamlit sebagai prototipe sistem pemantauan opini publik secara *real-time* bagi pemangku kepentingan (*stakeholders*).

4. KESIMPULAN

4.1 Kesimpulan

Penelitian ini telah berhasil mengonstruksi pemodelan klasifikasi sentimen otomatis dengan mengintegrasikan algoritma *Multinomial Naive Bayes* (MNB) dan teknik pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk memetakan persepsi publik terhadap digitalisasi pertanian di Indonesia. Berdasarkan implementasi kerangka kerja CRISP-DM, penerapan *relevancy filtering* berbasis kamus terbukti vital dalam menjamin validitas data melalui reduksi *noise* sebesar 64,63%, menghasilkan 1.156 komentar valid dari total 3.268 data mentah yang diekstraksi dari platform YouTube. Hasil pemetaan sentimen menunjukkan dominasi opini positif (292 data), yang mencerminkan optimisme serta keterbukaan masyarakat terhadap efisiensi teknologi *smart farming* dan penggunaan *drone*. Namun, keberadaan sentimen negatif yang signifikan (237 data) memberikan bukti empiris mengenai hambatan struktural di lapangan, khususnya tingginya modal awal pengadaan teknologi dan ketimpangan akses internet perdesaan. Evaluasi performa model menggunakan *Confusion Matrix* membuktikan tingkat keterandalan yang tinggi dengan pencapaian akurasi keseluruhan sebesar 83,39%. Secara keseluruhan, penelitian ini tidak hanya memperkaya literatur sains data dalam ranah pemrosesan bahasa Indonesia informal, tetapi juga memberikan instrumen *evidence-based policy* bagi pemerintah guna merumuskan strategi transformasi digital agrikultur yang lebih inklusif.

4.2 Saran

Guna meningkatkan kualitas generalisasi model di masa mendatang, disarankan bagi peneliti selanjutnya untuk melakukan ekspansi akuisisi data ke platform media sosial lain seperti X (Twitter), Instagram, maupun TikTok guna menangkap spektrum opini yang lebih heterogen. Selain itu, diperlukan pengayaan kuantitas data latih khususnya pada kategori netral serta integrasi modul normalisasi kata gaul (*slang*) yang lebih

komprehensif untuk meningkatkan sensitivitas klasifikasi. Pengembangan lebih lanjut juga dapat diarahkan pada eksplorasi arsitektur *Deep Learning* seperti BERT (*Bidirectional Encoder Representations from Transformers*) untuk menangkap makna kontekstual yang lebih kompleks, termasuk pendeteksian unsur sarkasme yang sering muncul dalam interaksi media sosial masyarakat Indonesia.

5. DAFTAR PUSTAKA

- Absharina, E. D. dan Nurqolbiah, F., 2025. Analisis Sentimen Komentar Video YouTube Review Starlink Indonesia Oleh Gadgetin Menggunakan VADER Dan TextBlob. *Journal of Applied Informatics*, 14(3), hlm. 4383–4389.
- Ardiansyah, A., Agustina, C., Maryani, I. dan Pribadi, D., 2025. Analisis Sentimen Pada Komentar YouTube Terkait Pembahasan ESIM Menggunakan Metode Naive Bayes Dan Random Forest. *Indonesian Journal on Software Engineering (IJSE)*, 11(1), hlm. 7–14. <https://doi.org/10.31294/ijse.v11i1.26180>.
- Fattah, Z. dan Syarifah, L., 2025. Analisis Sentimen Komentar YouTube Pada Video Terkait Insiden Pengemudi Ojek Online Dan Anggota Brimob Menggunakan Algoritma Naive. *Jurnal Mahasiswa Teknik Informatika*, 4(2), hlm. 90–96. <https://doi.org/10.35473/jamastika.v4i2.4495>.
- Harahap, S. M. dan Kurniawan, R., 2024. Analisis Sentimen Komentar YouTube Terhadap Food Vlogger Dengan Menggunakan Metode Naive Bayes. *MEANS (Media Informasi Analisa dan Sistem)*, 9(1), hlm. 87–96. <https://doi.org/10.54367/means.v9i1.3912>.
- Hidayat, T., Bahtiar, A. dan Kaslani, K., 2025. Classification of YouTube User Sentiment on 5G Technology Videos with Naive Bayes. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(3), hlm. 1706–1711. <https://doi.org/10.59934/jaiea.v4i3.992>.
- Kusuma, M. R., Gata, W., Kurniawati, L., Rivan, A. F. dan Kurniawan, T., 2023. Analisis Sentimen Masyarakat Terhadap Informasi Penerapan PPN Atas Renovasi Dan Membangun Rumah Sendiri Pada Media Sosial YouTube Dengan Metode Svm Dan Naive Bayes. *Jurnal Digital Teknologi Informasi*, 6(2), hlm. 65–75. <https://doi.org/10.32502/digital.v6i2.4758>.
- Ma'rufudin, M. dan Yudhistira, A., 2025. Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Pendidikan dan Teknologi Indonesia*, 5(3), hlm. 845–857. <https://doi.org/10.52436/1.jpti.717>.
- Maulida, S. dan Devi, A., 2023. Mining Netizen's Opinion on Sustainable Agriculture: Sentiment Analysis of Twitter Data. *Business and Sustainability*, 2(1). <https://doi.org/10.58968/bs.v2i1.323>.
- Nur, H. Y., Prasetyo, H., Wibowo, A. dan Kurniawan, D., 2025. Community Perspectives in the Implementation of Agricultural Digitalization. *Journal of Agricultural Transformation*, 2, hlm. 45–58.
- Pernama, T. D., Sutrisno, B., Haryanto, A. dan Kusuma, D., 2025. Sentiment Analysis of YouTube User Comments on Government Policies Using the Naive Bayes Method. *Journal Of Data Insights*, 3(2), hlm. 113–123. <https://doi.org/10.26714/jodi.v3i2.876>.
- Prabowo, T., Sarif, M. I., Sebayang, A., Ferdillah, T. D. dan Azuan, M. (2026) 'Analisis Sentimen Masyarakat terhadap Isu Korupsi Dana Bencana di Indonesia Menggunakan Metode Bidirectional Long Short-Term Memory (Bi-LSTM)', *Jurnal Komputer Teknologi Informasi Sistem Komputer (JUKTISI)*, 4(3), pp. 1787–1795. doi: 10.62712/juktisi.v4i3.756.
- Rizki, J., 2023. Social Media Sentiment Analysis to Understand Agricultural Market Trends and Consumer Preferences. *Jurnal Minfo Polgan*, 12(2), hlm. 1734–1743. <https://doi.org/10.33395/jmp.v12i2.12970>.
- Umaruddin, M., Febriansyah, A. dan Tahira, N., 2025. The Construction of Meaning and Sentiment: Agricultural Representation on YouTube in Indonesia. *Society*, 13(1), hlm. 369–387. <https://doi.org/10.33019/society.v13i1.804>.

