OPEN ACCESS

Volume 9 (2), April-June 2025, 485-494

E-ISSN:2580-1643

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: https://doi.org/10.35870/jtik.v9i2.3312

Analisis Sentimen Proyek Strategis Nasional *Food Estate* Menggunakan Algoritma *Naïve Bayes*, *Logistic Regression* dan *Support Vector Machine*

Yuning Rum Zattayu Mustopo [1*], Afiyati [2]

[1,2] Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta, Indonesia.*

a r t i c l e i n f o

Article history:
Received 15 October 2024
Received in revised form
10 November 2024
Accepted 20 December 2024
Available online April 2025.

Keywords:
Food Estate; Logistic Regression; Naïve Bayes; Sentiment Analysis; Support Vector Machine.

Kata Kunci:
Analisis Sentiment; Food Estate; Logistic Regression; Naïve Bayes; Support Vector Machine.

a b s t r a c t

The National Strategic Project Food Estate is an initiative by the Indonesian government aimed at enhancing food security through the development of large-scale agricultural areas. In the vice-presidential debate ahead of the 2024 election, Food Estate re-emerged as a hot topic, sparking controversy. Therefore, this study aims to analyze public perspectives on the National Strategic Project Food Estate by comparing the performance of machine learning algorithms, including Naïve Bayes, Logistic Regression and Support Vector Machine. This research also experiments with feature extraction techniques TF-IDF and Word2Vec. The results indicate that TF-IDF feature extraction performs better in capturing relevant features to enhance classification performance compared to the Word2Vec method. The best-performing algorithm is Logistic Regression + TF-IDF, achieving an accuracy of 74%, followed by SVM + TF-IDF and Naïve Bayes + TF-IDF with accuracies of 73% and 72%, respectively.

a b s t r a k

Proyek Strategis Nasional Food Estate merupakan program pemerintah Indonesia yang bertujuan untuk meningkatkan ketahanan pangan melalui pengembangan kawasan pertanian berskala besar. Pada debat cawapres menjelang pemilu 2024, Food Estate kembali menjadi topik hangat yang memunculkan kontroversi. Sehingga, penelitian ini bertujuan untuk menganalisis perspektif masyarakat mengenai Proyek Strategis Nasional Food Estate dengan membandingkan kinerja algoritma machine learning diantaranya yaitu Support Vector Machine, Naïve Bayes dan Logistic Regression. Kemudian penelitian ini juga melakukan percobaan dengan ektraksi fitur TF-IDF dan Word2Vec. Hasilnya ekstraksi fitur dengan TF-IDF lebih unggul dalam mengekstraksi fitur yang relevan untuk meningkatkan performa klasifikasi dibandingkan dengan metode Word2Vec. Algoritma dengan klasifikasi terbaik yaitu Logistic Regression+TF-IDF dengan akurasi sebesar 74%, selanjutnya yaitu Support Vector Machine+TF-IDF, Naïve Bayes+TF-IDF dengan akurasi sebesar 73% dan 72%.

ACM Computing Classification System (CCS)
EBSCOhost
Communication and Mass Media Complete (CMMC)

Corresponding Author. Email: 41520120067@student.mercubuana.ac.id [1].

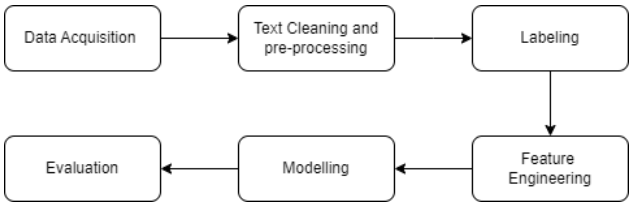
1. Pendahuluan

Konsep lumbung pangan skala besar, yang juga dikenal sebagai "*Food Estate*", adalah sebuah program dimana pemerintah memenuhi kebutuhan pangan lokal dengan memaksimalkan sektor swasta melalui alokasi investasi global dalam produksi pangan (Simamora *et al*., 2021). *Food Estate* adalah kawasan pertanian yang dikembangkan secara besar-besaran untuk menghasilkan pangan yang dibutuhkan masyarakat setempat. Pengembangan *Food Estate* diharapkan dapat membantu memenuhi kebutuhan pangan Indonesia untuk menghadapi krisis pangan di masa depan (Lasminingrat & Efriza, n.d.). Pada debat cawapres menjelang pemilu 2024, *Food Estate* kembali menjadi topik hangat yang memunculkan kontroversi, sehingga muncul beragam tanggapan dari para peserta debat dengan beragam sudut pandang yang memicu perbincangan yang mendalam di kalangan masyarakat. Proyek Strategis Nasional *Food Estate* ini telah di bahas dalam beberapa penelitian terutama dalam konteks sosial dan politik. Namun, dalam ruang lingkup teknologi informasi, terutama dalam survei kuantitatif penelitian terkait dengan topik ini masih belum ditemukan.

Hal tersebut dikemukakan oleh penelitian sebelumnya yang dilakukan oleh Iqbal yanuar (2024) mengenai "Sekuritisasi Dalam Kebijakan *Food Estate* Di Era Pemerintahan Joko Widodo" pada penelitian tersebut menyatakan bahwa belum ada data sekunder berupa survei kuantitatif yang menggambarkan persepsi masyarakat terhadap program *Food Estate*, yang menjadi sebuah celah penting dalam evaluasi kesuksesannya (Yanuar & Polsight, n.d.). Sehingga, hal tersebut menjadi ruang lingkup penelitian yang menarik untuk di eksplorasi lebih lanjut. Analisis sentimen adalah metode untuk mengklasifikasikan teks berdasarkan opini dan persepsi masyarakat. Proses ini melibatkan pemahaman dan pengolahan data teks secara otomatis untuk mengekstrak informasi berharga (Hasri & Alita, 2022). Pada penelitian ini analisis sentimen digunakan untuk mengklasifikasikan polaritas teks dan opini yang terkandung dalam teks dan tweet media sosial, apakah positif atau negatif terhadap persepsi masyarakat mengenai Proyek Strategis Nasional *Food Estate*. Penelitian yang dilakukan oleh Alvi Rahmy Royyan dan Erwin Budi Setiawan pada tahun 2022 mengenai analisis sentiment di Twitter terhadap kebijakan publik dengan metode ekspansi fitur *Word2Vec*. *Support Vector Machine* dan *Logistic Regression* menghasilkan akurasi yang cukup baik dengan masing-masing yaitu 78.99% dan 78.31%(Naufal & Setiawan, n.d.). Penelitian kedua yang dilakukan oleh Ahmad Hilman dkk berdasarkan hasil penelitian yang dilakukan, metode TF-IDF dengan SVM memberikan performa terbaik dalam analisis sentimen cyberbullying, dengan akurasi sebesar 84% dibandingkan dengan model berbasis *Word2Vec* (CBOW dan Skip-gram) dengan akurasi 77% dan 78% (Ahmad *et al*., 2024). Penelitian ketiga oleh Akhmad Muzaki dan Arita Witanti membahas analisis sentimen terkait PILKADA 2020 yang berlangsung di tengah pandemi COVID-19, menggunakan metode klasifikasi Naive Bayes dan fitur ekstraksi TF-IDF. Hasil penelitian ini menunjukkan bahwa akurasi terbaik yang dicapai adalah sebesar 92,2% (Muzaki & Witanti, 2021). Penelitian ini bertujuan untuk menganalisis persepsi masyarakat terhadap kebijakan pemerintah dalam Proyek Strategis Nasional *Food Estate* melalui analisis sentimen di media sosial. Penelitian ini akan mengklasifikasikan sentimen masyarakat menjadi dua kategori, yaitu positif dan negatif, dengan menggunakan tiga algoritma: *Naive Bayes*, *Logistic Regression* dan *Support Vector Machine*. Ekstraksi fitur menggunakan dua metode, yaitu TF-IDF dan *Word2Vec*, yang telah menunjukkan performa optimal dalam penelitian sebelumnya untuk berbagai topik di media sosial. Dengan membandingkan hasil dari kedua metode ekstraksi fitur ini, penelitian ini diharapkan dapat mengidentifikasi algoritma dan metode yang paling efektif dalam mengklasifikasikan sentimen publik terhadap Proyek Strategis *Food Estate*.

2. Metodologi Penelitian

Pendekatan penelitian ini menggunakan jenis penelitian kuantitatif. Algoritma *Naïve Bayes*, *Logistic Regression* dan *Support Vector Machine*, digunakan untuk mengevaluasi kinerja dari algoritma tersebut dengan membandingkan dan menentukkan algoritma mana yang paling efektif untuk digunakan dalam analisis sentiment yang dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

Data Acquisition

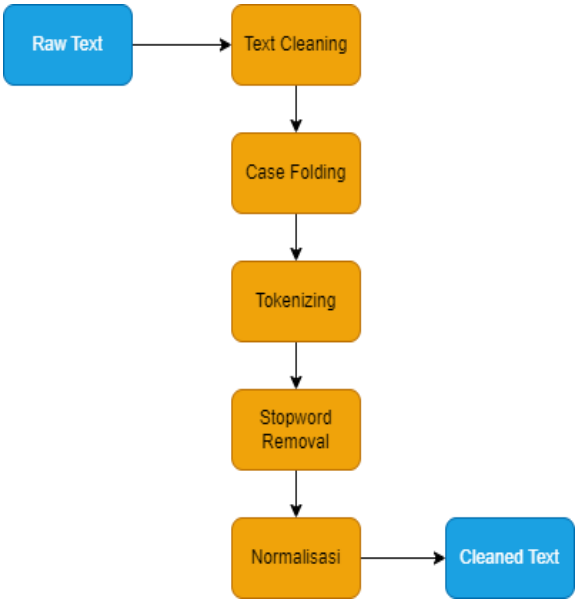
Data pada penelitian ini dikumpulkan dengan tools Tweet-Harvest dengan melakukan *crawling* data pada media sosial X atau Twitter untuk mendapatkan topik mengenai Proyek Strategis Nasional *Food Estate* dalam bentuk *tweet*/*thread* atau komentar. Kemudian tools ini dijalankan dengan Bahasa pemrograman python. Data yang dikumpulkan dicari menggunakan kata kunci "*food estate*" dari bulan Februari hingga Agustus, kemudian data *tweet* yang didapat diubah kedalam data frame menggunakan library pandas dan disimpan dalam format csv.

Text Cleaning and Pre-processing

Pembersihan teks dan *pre-processing* merupakan tahapan data wrangling dalam proyek *Natural Language Processing* (NLP). Yaitu proses mengekstrak teks mentah (*raw text*), atau teks asli, dari data input, menghilangkan informasi yang tidak penting, dan mengonversi teks ke dalam format yang dibutuhkan. Adapun tahapannya sebagai berikut dapat dilihat pada Gambar 2.

1) *Cleansing*
   Cleansing merupakan langkah untuk menghilangkan simbol atau karakter yang tidak diperlukan dalam dokumen.
2) *Case Folding*
   Case folding merupakan sebuah proses yang dilakukan untuk mengkonversi dokumen teks menjadi huruf kecil (Samantri, 2024).
3) *Stopword removal (filtering)*
   Stopword adalah kata-kata yang sering muncul dalam teks, namun tidak memiliki nilai informatif. yang tinggi seperti "dan", "di", "ke", "yang", dan sebagainya (Safitri *et al.*, 2023).
4) *Tokenizing*
   Proses tokenisasi melibatkan pemisahan teks menjadi bagian-bagian kecil yang disebut token (Suryati *et al.*, 2023). Tokenisasi digunakan untuk memecah teks menjadi bagian-bagian yang lebih mudah diolah atau ditafsirkan melalui analisis teks atau pemrosesan bahasa alami.
5) *Normalisasi*
   Proses ini dilakukan untuk memastikan bahwa kata-kata yang diperpanjang atau disingkat disesuaikan dengan kata-kata yang normal sesuai dengan definisi yang ditemukan dalam Kamus Besar Bahasa Indonesia (KBBI). Mengubah kata slang menjadi kata baku disebut konversi slang *word* (Gifari *et al.*, 2022).

Pre-processing bertujuan untuk membersihkan, menormalkan, dan mengubah data mentah menjadi format yang lebih sederhana dan mudah dipahami oleh algoritma. Proses ini membantu untuk mengekstrak informasi penting yang dapat meningkatkan akurasi dalam proses klasifikasi data.



Gambar 2. Tahap *Preprocessing*

Labeling

Pelabelan data pada penelitian ini menggunakan kamus *Inset Lexicon* Indonesia. Kamus ini terdiri dari 3.609 kata positif dan 6.609 kata negatif dengan skor berkisar -5 dan +5. Metode pelabelan ini menggunakan bobot setiap kata dalam kamus untuk menentukan apakah kalimat tersebut memiliki sentimen positif atau negatif. Untuk menghitung nilai rata-rata sentimen, jumlah bobot setiap kata dibagi dengan jumlah kata dalam kalimat. Kalimat itu dinilai negatif atau positif berdasarkan hasil rata-ratanya.

Data Splitting

Pembagian dataset pada penelitian ini yaitu dengan perbandingan 90:10, 5871 untuk data *training* dan 542 untuk data test. Atribut yang digunakan yaitu "*full_text*" (teks lengkap dari *tweet*) kemudian dibagi menjadi empat variabel (*x_train, y_train, x_test, dan y_test*) menggunakan fungsi "*train_test_split*" dari modul "*sklearn model selection*" dalam *library sklearn*.

Feature Engineering

Proses pembobotan setiap kata dikenal sebagai *term weighting* yaitu sebuah proses untuk mengoptimalkan kemampuan analisis sentimen dalam proses *text mining* (Gifari *et al*., 2022). TF-IDF dilakukan dengan mengalikan dua metrik yang dapat diperesentasikan sebagai tf - idf = tf * idf, dimana tf (*term frequency*) merupakan frekuensi kemunculan dari suatu term/istilah dan idf (*inverse document frequency*) merupakan frekuensi inverse document dari term/istilah tersebut. Dalam TF-IDF, frekuensi sebuah kata dalam dokumen dihitung lalu dikompensasikan dengan jumlah dokumen yang memiliki kata itu. Selain itu penelitian ini akan menggunakan pembobotan kata dengan *Word2Vec*. Word2Vec adalah algoritma vektor kata yang memberikan performa optimal di NLP dengan mengelompokkan kata-kata serupa ke dalam vektor yang mirip (Nawangsari *et al*., 2019). Word2Vec memanfaatkan Jaringan Syaraf Tiruan untuk memproses teks sebagai input dan menghasilkan representasi dalam bentuk ruang vektor. Hasil vektor kata berbentuk ruang berdimensi rendah yang mampu menangkap arti semantik dari kata yang direpresentasikan.

Modeling

Setelah melakukan *feature engineering* dengan tf-idf, langkah selanjutnya yaitu pemodelan dengan algoritma *Support Vector Machine*, *Naïve Bayes* dan *Logistic Regression* menggunakan empat skenario data *training* dan data *testing* dengan perbandingan 90:10 untuk mendapatkan akurasi dari masing-masing algoritma dari skenario yang telah dibuat. Kemudian, hasil dari pemodelan tersebut dibandingkan menggunakan metrik akurasi.

Evaluation

Hasil penelitian ini dievaluasi menggunakan *confusion matrix* untuk mengukur kinerja algoritma machine learning yang digunakan. Confusion matrix adalah matriks yang menyajikan data klasifikasi aktual dan prediksi, yang digunakan untuk mengevaluasi kinerja model klasifikasi (Homepage *et al*., 2021). *Tabel confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

		Kelas Aktual	
		Positive	*Negative*
Kelas Prediksi	*Positive*	*True Positive* (TP)	*False Positive (FP)*
	Negative	*False Negative* (TN)	*Tru e Negative* (TN)

Pada penelitian ini performa klasifikasi yang akan dihitung adalah *accuracy, precision, recall* dan *f1 score*. Adapun rumus-rumus yang digunakan dapat dilihat pada persamaan (1), (2), (3), dan (4).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$precission = \frac{TP}{TP+FP}$$

$$f1-score = 2\text{X}\frac{recall \times precission}{recall + precission}$$

3. Hasil dan Pembahasan

Hasil

Dataset

Dataset dikumpulkan melalui proses *crawling* data pada media sosial X atau Twitter, *tweet* dikumpulkan menggunakan Twitter API menggunakan Bahasa pemrograman *Python*, berupa teks yang mengandung kata kunci "*food estate*" yang diambil dari rentang waktu Februari 2024 sampai Agustus 2024 sebanyak 5530 data. *Dataset* dapat dilihat pada gambar 3.

	full_text	username	created_at
0	@Rachel_Lonie Jgn bikin orang susah katanya.....	NaufalAriiq26	Fri Aug 09 23:51:15 +0000 2024
1	@hidahidaan Mbok ya anda itu jangan asal cuap ...	abi_daawud	Fri Aug 09 13:10:02 +0000 2024
2	@erlanishere halo siri apa relasi food estate ...	disor_	Fri Aug 09 12:58:54 +0000 2024
3	Tingkatkan Minat Baca Bhabinkamtibmas Food Est...	aping_wae	Fri Aug 09 12:27:08 +0000 2024
4	@SahalaO82038 @BANGSAygSUJUD @prabowo @99propa...	DarekaRendy	Fri Aug 09 11:47:37 +0000 2024
...	...	...	...
5525	@tanyarlfes ya bisa bisa aja tpi minimal akses...	baxxxoxxxapi___	Sat Mar 02 07:05:35 +0000 2024
5526	@matchaacheese_2 @ikbal_fari34568 @Kacaback678...	LFHLEON	Sat Mar 02 07:04:20 +0000 2024
5527	@tvOneNews @AgusYudhoyono ok saya dukung tapi ...	sabang_marouke	Sat Mar 02 07:00:03 +0000 2024
5528	@mas2_jawa_ @ARSIPAJA Ga bisa lah bang. Soalny...	blvdhc__	Sat Mar 02 06:54:03 +0000 2024
5529	@PartiBurungGalak Food estate bagus di ide ga b...	herrycho_12	Sat Mar 02 06:52:42 +0000 2024

5530 rows × 3 columns

Gambar 3. Dataset

Preprocessing

Preprocessing ini dilakukan dengan tujuan untuk Membersihkan data dari noise, mengubah teks ke dalam format yang konsisten, mengekstrak fitur-fitur penting untuk analisis lanjutan dan meningkatkan akurasi. Adapun tahapannya sebagai berikut:

1) *Cleansing*

Tahap *cleansing* yaitu membersihkan teks dari angka, simbol, tanda baca, atau karakter-karakter khusus yang tidak relevan, seperti terlihat pada tabel 2.

Tabel 2. Tahap *Cleansing*

Sebelum	Sesudah
Ngomong yg realistis ajalah drpd thn lg sibuk ngeles Saat ini aja sy dah capek liat orang di sekitar anda berdebat di layar kaca Lebih baik bicara proyek food estate itu mo diapain Jujur aja sy gak kebayang Indonesia ini kedepannya akan seperti apa	Ngomong yg realistis ajalah drpd thn lg sibuk ngeles Saat ini aja sy dah capek liat orang di sekitar anda berdebat di layar kaca Lebih baik bicara proyek food estate itu mo diapain Jujur aja sy gak kebayang Indonesia ini kedepannya akan seperti apa

2) *Case Folding*

Pada tahap *Case Folding,* semua huruf dalam dokumen teks dikonversi menjadi huruf kecil, seperti terlihat pada tabel 3.

Tabel 3. Tahap *Case Folding*

Sebelum	Sesudah
Ngomong yg realistis ajalah drpd thn lg sibuk ngeles Saat ini aja sy dah capek liat orang di sekitar anda berdebat di layar kaca Lebih baik bicara proyek food estate itu mo diapain Jujur aja sy gak kebayang Indonesia ini kedepannya akan seperti apa	ngomong yg realistis ajalah drpd thn lg sibuk ngeles saat ini aja sy dah capek liat orang di sekitar anda berdebat di layar kaca lebih baik bicara proyek food estate itu mo diapain jujur aja sy gak kebayang indonesia ini kedepannya akan seperti apa

3) *Normalisasi*

Proses normalisasi yaitu menyederhanakan teks dengan mengubah istilah slang menjadi kata-kata baku sesuai dengan kaidah bahasa, seperti terlihat pada tabel 4.

Tabel 4. Tahap Normalisasi

Sebelum	Sesudah
ngomong yg realistis ajalah drpd thn lg sibuk ngeles saat ini aja sy dah capek liat orang di sekitar anda berdebat di layar kaca lebih baik bicara proyek food estate itu mo diapain jujur aja sy gak kebayang indonesia ini kedepannya akan seperti apa	bicara yang realistis ajalah daripada thn lagi sibuk banyak alasan saat ini aja saya sudah lelah liat orang di sekitar anda berdebat di layar kaca lebih baik bicara proyek food estate itu mau diapain jujur aja saya tidak membayangkan indonesia ini kedepannya akan seperti apa

4) *Tokenizing*

Proses *tokenizing* yaitu proses membagi teks menjadi unit-unit yang lebih mudah untuk diproses atau diinterpretasikan dalam analisis teks. Unit-unit ini, atau token, dapat berupa kata, frasa, atau elemen lain yang lebih kecil dari teks yang dianalisis, seperti terlihat pada tabel 5.

Tabel 5. Tahap *Tokenizing*

Sebelum	Sesudah
bicara yang realistis ajalah daripada thn lagi sibuk banyak alasan saat ini aja saya sudah lelah liat orang di sekitar anda berdebat di layar kaca lebih baik bicara	['bicara', 'yang', 'realistis', 'ajalah', 'daripada', 'thn', 'lagi', 'sibuk', 'banyak', 'alasan', 'saat', 'ini', 'aja', 'saya', 'sudah', 'lelah', 'liat', 'orang', 'di', 'sekitar', 'anda', 'berdebat', 'di', 'layar',

proyek food estate itu mau diapain jujur aja saya tidak membayangkan indonesia ini kedepannya akan seperti apa	'kaca', 'lebih', 'baik', 'bicara', 'proyek', 'food', 'estate', 'itu', 'mau', 'diapain', 'jujur', 'aja', 'saya', 'tidak', 'membayangkan', 'indonesia', 'ini', 'kedepannya', 'akan', 'seperti', 'apa']

5) *Stopword Removal (Filtering)*
Proses penghapusan *stopword* ini yaitu membersihkan teks dari kata-kata umum, sehingga analisis dapat lebih terfokus pada kata-kata kunci yang lebih bermakna, seperti terlihat pada tabel 6.

Tabel 6. Tahap *Stopword Removal*

Sebelum	Sesudah
['bicara', 'yang', 'realistis', 'ajalah', 'daripada', 'thn', 'lagi', 'sibuk', 'banyak', 'alasan', 'saat', 'ini', 'aja', 'saya', 'sudah', 'lelah', 'berdebat', 'layar', 'liat', 'orang', 'di', 'sekitar', 'anda', 'berdebat', 'di', 'layar', 'kaca', 'lebih', 'baik', 'bicara', 'proyek', 'food', 'estate', 'itu', 'mau', 'diapain', 'jujur', 'aja', 'saya', 'tidak', 'membayangkan', 'indonesia', 'ini', 'kedepannya', 'akan', 'seperti', 'apa']	['bicara', 'realistis', 'ajalah', 'thn', 'sibuk', 'alasan', 'aja', 'lelah', 'liat', 'orang', 'kaca', 'bicara', 'proyek', 'food', 'estate', 'diapain', 'jujur', 'aja', 'membayangkan', 'indonesia', 'kedepannya']

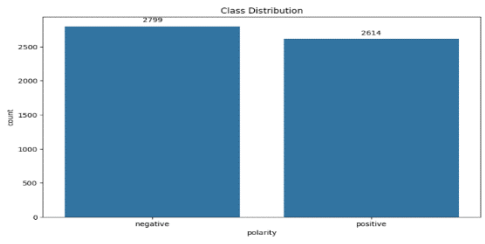
Labeling

Pelabelan pada penelitian ini menggunakan kamus Indonesia *Inset Lexicon*, kamus ini terdiri dari 2 daftar kata yaitu positif.tsv dan negatif.tsv, Setelah itu, setiap baris dalam file dibaca dan dimasukkan ke dalam kamus yang sesuai (*lexicon*_positif untuk kata-kata positif dan *lexicon*_negative untuk kata-kata negatif). Contoh hasil pelabelan pada dataset dapat dilihat pada tabel 7.

Tabel 7. Contoh Hasil Pelabelan

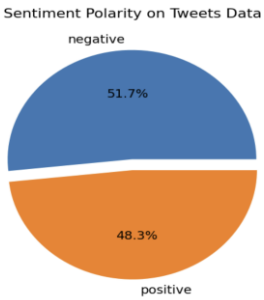
Full text	*polarity_score*	*polarity*
bikin krisis pangan tuh *food estate* gagal manajemen logistik pangan buruk….	-9	*negative*
food estate gagal utang menggunung lingkungan hidup hancur demokrasi hancur bangsa….	-10	*negative*
food estate mandiri pemerintah alhamdulillah panen.	2	*positive*
program *food estate* temanggung berhasil petani rasakan manfaatnya…	6	*positive*

Dataset terdiri dari 5.413 *tweet* yang mengandung informasi mengenai *food estate*. Setelah proses pelabelan, diperoleh dataset dengan *sentiment positive* sebanyak 2.614 data dan *sentiment negative* sebanyak 2.799 data, sehingga *class distribution* dapat dilihat pada gambar 4.



Gambar 4. Distribusi Kelas

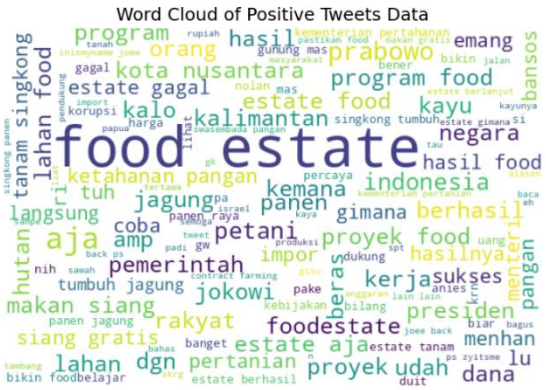
Polaritas sentimen dari dataset yang digunakan terlihat *sentiment positive* lebih kecil yaitu 48.3% dibandingkan dengan *sentiment negative* yaitu sebesar 51.7%, seperti yang terlihat pada gambar 5.



Gambar 5. Presentase Kelas

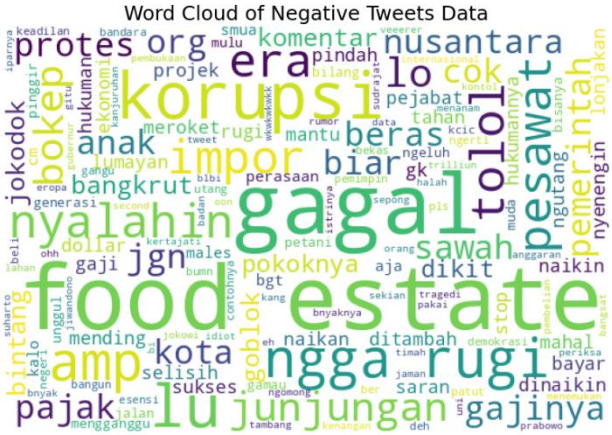
Visualisasi Data

Visualisasi data dalam penelitian ini bertujuan untuk mempermudah pemahaman dan interpretasi data, serta mengidentifikasi temuan-temuan penting di dalamnya. Visualisasi tersebut disajikan dalam bentuk bar plot dan *wordcloud*. Terdapat dua kategori utama dari distribusi kelas tersebut yaitu sentiment positif dan 491ocus491ve yang menunjukkan bahwa jumlah sentiment negative sedikit lebih tinggi dari sentiment positif. Meskipun tidak terlalu signifikan, hal ini mengindikasikan bahwa kecendrungan respon 491ocus491ve lebih besar terhadap Proyek Strategis Nasional *Food Estate*. Kecendrungan tersebut mencerminkan bahwa dalam implementasinya Proyek Strategis Nasional *Food Estate* ini masih menuai banyak kritikan dan ketidakpuasan dalam kalangan Masyarakat yang perlu mendapat perhatian lebih lanjut dalam evaluasi kebijakan dan pelaksanaan proyek. *WordCloud* adalah representasi visual dari kata-kata yang muncul dalam teks, ketika ukurannya semakin besar menunjukkan bahwa seberapa sering kata tersebut muncul. Pada Gambar 6. Kata "*food estate*" menjadi yang paling dominan, menunjukkan 491ocus utama diskusi. Kata-kata lain seperti "program", "berhasil", "petani", "pemerintah", dan "panen" juga muncul cukup besar, mengindikasikan bahwa banyak *tweet* positif yang mendukung proyek.



Gambar 6. *Wordcloud* Kata Positif

Pada gambar 7 *wordcloud* negatif kata-kata yang paling sering muncul dalam tweet. Kata "*food estate*" tetap dominan, namun kata-kata negatif seperti "gagal", "korupsi", "rugi", dan "impor" juga terlihat besar, mengindikasikan bahwa banyak sentimen negatif terkait dengan pandangan proyek tersebut sebagai kegagalan, adanya potensi korupsi, kerugian ekonomi, atau kebijakan impor yang tidak disukai.



Gambar 7. *Wordcloud* Kata Negatif

Feature Engineering dan Data Splitting

Penelitian ini menggunakan pembobotan kata dengan TF-IDF, data dipisahkan menjadi dua bagian, yaitu fitur (*tweet*) yang diwakili oleh variabel X dan label sentimen yang diwakili oleh variabel y. Fitur *tweet* diekstraksi menggunakan metode TF-IDF. Dalam proses TF-IDF, pengaturan spesifik yang digunakan meliputi jumlah fitur maksimum (*max_features*), frekuensi minimum kata (min_df), dan persentase maksimum dokumen yang mengandung kata tersebut (max_df). Hasil ekstraksi fitur tersebut kemudian diubah menjadi *DataFrame* untuk mempermudah analisis lanjutan. Selain dibobotkan dengan menggunakan TF-IDF, data yang telah dibagi menjadi data uji dan data latih juga akan diberi bobot menggunakan *Word2Vec* dengan bantuan *library gensim*. Pembagian data sebanyak 90:10 untuk data *training* dan data *testing*.

Pengujian dan Evaluasi

Setelah melakukan ekstraksi fitur dengan TF-IDF dan *Word2Vec*, pemodelan dilakukan menggunakan ketiga algoritma tersebut (*Naïve Bayes*, dan *Logistic Regression* dan *Support Vector Machine*). Hasil dari pemodelan ini kemudian dievaluasi menggunakan *confusion matrix*, yang menunjukkan performa dari masing-masing algoritma dan ekstraksi fitur dalam mengklasifikasikan data. Hasil evaluasi dengan *confusion matrix* untuk setiap algoritma dapat dilihat pada tabel 8.

Tabel 8. Hasil dengan TF-IDF

Algoritma	Sentiment	*Accuracy*	*Precission*	*Recall*	*F1-Score*
SVM	Positif	0.73	0.70	0.79	0.74
	Negatif		0.86	0.67	0.71
Naive Bayes	Positif	0.72	0.70	0.77	0.73
	Negatif		0.74	0.67	0.71
Logistic Regression	Positif	0.74	0.73	0.77	0.75
	Negatif		0.76	0.72	0.74

Berdasarkan hasil evaluasi yang dilakukan, dapat disimpulkan bahwa metode ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) memberikan performa yang lebih baik dibandingkan metode *Word2Vec* dalam konteks pemodelan menggunakan algoritma *Naïve Bayes*, *Logistic Regression* dan *Support Vector Machine*. Hal ini dibuktikan dengan nilai akurasi yang lebih tinggi ketika menggunakan TF-IDF, di mana masing-masing algoritma SVM, *Naïve Bayes*, dan *Logistic Regression* mencapai akurasi sebesar 73%, 72%, dan 74%. Sebaliknya, metode *Word2Vec* hanya menghasilkan akurasi sebesar 60%, 59%, dan 58% pada model yang sama. Hasil ini menunjukkan bahwa dalam konteks analisis sentiment Proyek Strategis Nasional *Food Esstate*, metode TF-IDF lebih unggul dalam mengekstraksi fitur yang relevan untuk meningkatkan performa klasifikasi dibandingkan dengan metode *Word2Vec*, seperti terlihat pada tabel 9.

Tabel 9. Hasil dengan *Word2Vec*

Algoritma	Sentiment	*Accuracy*	*Precission*	*Recall*	*F1-Score*
SVM	Positif	0.60	0.70	0.33	0.45
	Negatif		0.56	0.86	0.68
Naive Bayes	Positif	0.59	0.56	0.37	0.47
	Negatif		0.65	0.80	0.66
Logistic Regression	Positif	0.58	0.62	0.41	0.49
	Negatif		0.56	0.75	0.64

Pembahasan

Dalam penelitian ini menunjukkan bahwa persepsi masyarakat terhadap Proyek Strategis Nasional *Food Estate* cenderung lebih negatif, meskipun terdapat harapan positif terkait ketahanan pangan. Berdasarkan hasil analisis sentimen yang dilakukan dengan menggunakan tiga algoritma *Naive Bayes*, *Logistic Regression*, dan *Support Vector Machine* (SVM) dengan dua metode ekstraksi fitur, *TF-IDF* dan *Word2Vec*, ditemukan bahwa sebagian besar sentimen yang diungkapkan oleh masyarakat di media sosial berfokus pada kekhawatiran terkait dampak sosial-ekonomi dan transparansi kebijakan. Hal ini sejalan dengan penelitian sebelumnya oleh Lasminingrat & Efriza (n.d.), yang menunjukkan bahwa masyarakat sering kali memandang proyek ini dengan skeptisisme, terutama terkait dengan keberlanjutannya di masa depan. Hasil analisis dengan algoritma *Naive Bayes* dan *Logistic Regression* menunjukkan dominasi sentimen negatif, yang berhubungan dengan persepsi buruk terhadap dampak jangka panjang dari *Food Estate* terhadap masyarakat lokal. Sebaliknya, algoritma SVM memberikan hasil yang lebih seimbang, mampu mengidentifikasi baik sentimen positif maupun negatif dengan lebih baik, menunjukkan kemampuannya dalam menangani data yang lebih kompleks dan bervariasi. Dalam hal ekstraksi fitur, metode *TF-IDF* menunjukkan performa yang lebih baik dibandingkan dengan *Word2Vec*. *TF-IDF* terbukti lebih efektif dalam mengidentifikasi kata-kata penting berdasarkan frekuensi dan relevansinya dalam teks, yang lebih sesuai dengan karakteristik data yang lebih sederhana. Hal ini sejalan dengan temuan Nawangsari *et al.* (2019) yang menunjukkan bahwa *TF-IDF* lebih unggul dalam analisis sentimen terhadap

data yang tidak terlalu bergantung pada hubungan semantik antar kata. Sementara itu, meskipun *Word2Vec* menawarkan representasi kata yang lebih mendalam melalui embedding vektor, hasilnya tidak memberikan peningkatan signifikan pada analisis ini. Penggunaan metode *Word2Vec*, yang lebih sering menunjukkan hasil optimal pada teks yang lebih kompleks, tidak memberikan hasil yang lebih baik dalam konteks analisis sentimen terhadap *Food Estate*, sebagaimana terlihat pada penelitian ini dan juga penelitian oleh Naufal & Setiawan (n.d.) terkait penggunaan *Word2Vec* dalam analisis sentimen di *Twitter*. Temuan ini menunjukkan bahwa pemerintah perlu mengambil langkah-langkah lebih transparan dalam menyampaikan tujuan dan manfaat dari proyek *Food Estate*.

Berdasarkan penelitian oleh Simamora *et al.* (2021), komunikasi yang lebih terbuka dan partisipatif dapat membantu mengurangi ketidakpercayaan masyarakat terhadap proyek ini. Selain itu, pemerintah dapat memanfaatkan hasil analisis sentimen sebagai alat untuk merancang kebijakan yang lebih responsif terhadap aspirasi dan kekhawatiran publik. Penelitian ini juga menyarankan agar pemerintah menggunakan media sosial sebagai saluran untuk melakukan dialog langsung dengan masyarakat, memperbaiki pemahaman mereka mengenai *Food Estate*, serta mengklarifikasi isu-isu yang berkembang. Namun, penelitian ini juga memiliki keterbatasan, seperti terbatasnya jumlah data yang diambil dari media sosial, yang dapat mempengaruhi representativitas hasil. Penelitian lebih lanjut dengan menggunakan data yang lebih beragam dan mempertimbangkan faktor sosial dan politik yang lebih luas dapat memberikan gambaran yang lebih komprehensif mengenai persepsi publik terhadap proyek ini. Penggunaan teknologi analisis sentimen yang lebih canggih, seperti model *deep learning* (*BERT* atau *GPT*), juga dapat meningkatkan akurasi analisis untuk mengidentifikasi pola sentimen yang lebih halus dan kompleks.

4. Kesimpulan

Kesimpulan penelitian ini menunjukkan bahwa pandangan masyarakat terhadap Proyek Strategis Nasional *Food Estate* cenderung negatif, berdasarkan analisis sentimen dari 5.413 data, dengan sentimen negatif sebanyak 2799 dan sentimen positif sebanyak 2614. Selain itu, penelitian ini juga melakukan evaluasi kinerja tiga algoritma klasifikasi *Naïve Bayes*, *Logistic Regression* dan *Support Vector Machine* dengan dua metode ekstraksi fitur, yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Word2Vec*. Hasil pengujian menunjukkan bahwa TF-IDF secara konsisten menghasilkan akurasi yang lebih tinggi dibandingkan *Word2Vec*, dengan masing-masing algoritma mencapai akurasi 73% (*Support Vector Machine*), 72% (*Naïve Bayes*), dan 74% (*Logistic Regression*) saat menggunakan TF-IDF, dibandingkan dengan akurasi yang lebih rendah pada *Word2Vec* yaitu 60%, 59%, dan 58%. Dapat disimpulkan bahwa metode TF-IDF dengan algortima *Logistic Regression* lebih efektif dengan akurasi paling baik yaitu 74% dalam memodelkan opini publik terkait proyek ini. Temuan ini diharapkan dapat memberikan wawasan yang bermanfaat bagi pemerintah dalam memahami persepsi publik terhadap Proyek Strategis Nasional *Food Estate*, serta menjadi dasar dalam mengambil langkah-langkah yang lebih tepat untuk program tersebut di masa mendatang.

5. Daftar Pustaka

Anwarudin, M. A. (2024). *Analisis sentimen publik di media sosial Twitter terhadap program Food Estate menggunakan algoritma Support Vector Machine dan Naive Bayes Classifier* (Doctoral dissertation, UIN Sunan Gunung Djati Bandung).

Dani, A. H., Puspaningrum, E. Y., & Mumpuni, R. (2024). Studi Performa TF-IDF dan Word2Vec Pada Analisis Sentimen Cyberbullying. *Router: Jurnal Teknik Informatika dan Terapan*, *2*(2), 94-106. https://doi.org/10.62951/router.v2i2.76.

Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, *2*(1), 36-40. https://doi.org/10.46229/jifotech.v2i1.330.

Hasri, C. F., & Alita, D. (2022). Penerapan Metode Naïve Bayes Classifier Dan Support Vector

Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter. *Jurnal informatika dan rekayasa perangkat lunak*, *3*(2), 145-160.

Lasminingrat, L., & Efriza, E. (2020). The development of national food estate: The Indonesian food crisis anticipation strategy. *Jurnal Pertahanan dan Bela Negara*, *10*(3), 229-248.

Muzaki, A., & Witanti, A. (2021). Sentiment analysis of the community in the twitter to the 2020 election in pandemic covid-19 by method naive bayes classifier. *Jurnal Teknik Informatika (Jutif)*, *2*(2), 101-107.

Naufal, H. F., & Setiawan, E. B. (2021). Ekspansi Fitur Pada Analisis Sentimen Twitter Dengan Pendekatan Metode Word2Vec. *eProceedings of Engineering*, *8*(5).

Nawangsari, R. P., Kusumaningrum, R., & Wibowo, A. (2019). Word2vec for Indonesian sentiment analysis towards hotel reviews: An evaluation study. *Procedia Computer Science*, *157*, 360-366. https://doi.org/10.1016/j.procs.2019.08.178.

Purba, M., Ermatita, E., Abdiansah, A., Noprisson, H., Ayumi, V., Salamah, U., ... & Yadi, Y. (2022). Effect of Random Splitting and Cross Validation for Indonesian Opinion Mining using Machine Learning Approach. *Int. J. Adv. Comput. Sci. Appl*, *13*(9).

Ramadhan, I. Y. (2023). Sekuritisasi Dalam Kebijakan Food Estate di Era Pemerintahan Joko Widodo. *Aliansi: Jurnal Politik, Keamanan Dan Hubungan Internasional*, *2*(3), 153-163.

Safitri, T., Umaidah, Y., & Maulana, I. (2023). Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine. *Journal of Applied Informatics and Computing*, *7*(1), 34-41.

Samantri, M. (2024). Perbandingan Algoritma Support Vector Machine dan Random Forest untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *8*(1), 1-9. https://doi.org/10.35870/jti.

Simamora, B., Lubis, K., & Arini, H. (2021). Analisis asumsi-asumsi pada program food estate di Papua. *PERSPEKTIF*, *10*(2), 293-300. https://doi.org/10.31289/perspektif.v10i2.4267.

Suryati, E., Styawati, S., & Aldino, A. A. (2023). Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi dan Sistem Informasi*, *4*(1), 96-106. *4*(1), 96–106. https://doi.org/10.33365/jtsi.v4i1.2445.