TRANSFORMASI KABUPATEN BATANG SEBAGAI KOTA INDUSTRI : MENGANALISIS SENTIMEN DAMPAK PEMBANGUNAN KAWASAN INDUSTRI TERPADU BATANG MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Mita Nur Istiyani[1*], Khopsa Chanda Syaputri[2] , Amalia Putri Setya Aryana[3], Drajat Stiawan[4]
[1234]Sains Data,Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan, Indonesia

**Penulis korespondensi: mitanuristiyani@gmail.com*

ABSTRAK

Pembangunan Kawasan Industri Terpadu Batang merupakan salah satu proyek besar dari Proyek Strategis Nasional (PSN) yang bertujuan untuk mendorong pertumbuhan ekonomi dan menciptakan lapangan kerja baru. Penelitian ini dilakukan untuk menetahui opini masyarkat terhadap pembangunan proyek tersebut, dengan metode analisis sentimen dari komentar yang ada di beberapa postingan Instagram. Komentar dikumpulkan secara manual dari unggahan Instagram berhastag dari *#kitbatang, #grandbatangcity, #kitb*, kemudian dianalisis menggunakan algoritma *Support Vector Machine* (SVM) dengan dukungan metode TF-IDF untuk ekstraksi fitur teks. Proses analisis dilakukan secara bertahap melalui *preprocessing*, *exploratory data analysis* (EDA), *feature engineering*, dan evaluasi model. Hasil klasifikasi menunjukan 3,7% komentar positif dan 96,3% komentar negatif dari total 132 data. Model mencapai akurasi 96,3%, presisi 92,7%, dan *recall* 96,3%. Hal ini menunjukkan bahwa model sangat baik dalam mendeteksi komentar negatif namun gagal dalam mendeteksi beberapa komentar positif dikarenakan tidak keseimbangan data. Temuan ini menunjukkan bahwa meskipun banyaknya sentimen negatif dengan topik akses kerja serta dampak dari pembangunan, sentimen positif yang mendukung adanya pembangunan Kawasan Industri Terpadu Batang masih ada. Penelitian memiliki kontribusi dalam literatur analisis sentimen pembangunan melalui studi kasus di KIT Batang dengan machine learning menggunakan algoritma *Support Vektor Machine*, dimana penelitian ini mampu menangkap beberapa opini masyarakat dari komentar instagram secara *real-time.*

Kata kunci: Kawasan Industri Terpadu Batang, analisis sentimen, media sosial, Instagram, Support Vector Machine (SVM).

1 PENDAHULUAN

Pembangunan Industri menjadi salah satu strategi pemerintah untuk bisa meningkatkan perekonomian, penciptaan lapangan kerja, dan meningkatkan kualitas sumber daya pekerja untuk bisa bersaing global. Batang menjadi salah satu lokasi yang terpilih untuk pembangunan kawasan Proyek Strategis Nasional (PSN) berdasarkan Peraturan Presiden Republik Indonesia Nomor 109 Tahun(Industri et al., 2023). Pembangun Kawasan Industri Terpadu sudah berjalan selama 3 tahun hingga kini masih terus berlanjut sehingga menyebabkan perubahan tutupan lahan dan berbagai dampak lainnya. Proyek ini bertujuan untuk mendorong pertumbuhan ekonomi di Kabupaten Batang dengan menciptakan lapangan kerja dan menarik investasi(Meylinda, 2021). Di Kabupaten Kendal, pembangunan industri kerap kali memunculkan dinamika sosial baru seperti pergeseran mata pencaharian dan ketimpangan ekonomi antarwilayah (Kusumawati, 2023). Dampak pembangunan yang mempengaruhi penggunaan lahan, perubahan ekosistem, dan konsekuensi sosial di masyarakat sekitar. distribusi akses kerja dalam proyek strategis nasional cenderung tidak merata dan berpotensi menimbulkan kecemburuan sosial di tingkat lokal.

Seiring dengan perkembangan pembangunan KIT Batang, berbagai argumen muncul dari berbagai kelompok. Banyak pihak yang mendukung proyek ini, dengan harapan proyek ini akan membawa kemakmuran dan pembangunan ekonomi yang signifikan. Namun, tidak sedikit pula yang menolak pembangunan ini karena kekhawatiran terhadap dampak sosial dan lingkungan yang mungkin timbul, seperti peningkatan polusi udara dan air, serta perubahan struktur sosial yang dapat mempengaruhi kehidupan masyarakat. Ketegangan antara dampak positif dan negatif inilah yang menjadi fokus utama dalam diskusi mengenai masa depan Batang sebagai pusat industri.

Kawasan Industri Terpadu Batang dibangun pada lokasi yang strategis yaitu di jalur utara pulau Jawa. KIT Batang berjarak 50 km dengan Bandara Ahmad Yani Semarang, berdekatan dengan Kawasan Industri Kendal dan Pelabuhan Kendal, hal tersebut yang menjadikan lokasi ini lokasi yang mudah diakses(Industri et al., 2023). Seiring dengan pembangunan kawasan ini, berbagai infrastruktur seperti jalan, penyediaan air bersih, pengelolaan limbah, listrik, telekomunikasi, dan lain-lain juga mengalami perkembangan yang besar. Pembangunan infrastruktur ini tentu membawa dampak positif, terutama dalam mempercepat pertumbuhan industri dan ekonomi lokal. Namun, perubahan besar ini juga memunculkan tantangan baru, seperti kemacetan lalu lintas, beban tambahan pada sistem lingkungan, dan ketimpangan dalam pemerataan manfaat pembangunan bagi masyarakat sekitar. Permasalahan yang diangkat dalam penelitian ini adalah tentang apa saja dampak yang diberikan dengan adanya kawasan Industri Terpadu Batang. Berbagai penelitian sebelumnya sudah banyak yang mengkaji tentang dampak dari pembangunan industri terhadap pertumbuhan ekonomi, penciptaan lapangan pekerjaan, serta perubahan beberapa struktural masyarakat. Dalam penelitian ini memiliki posisi pembeda dari beberapa penelitian lainnya,karena berfokus kepada Kawasan Industri Terpadu Batang, yang saat ini masih dalam tahap pembangunan aktif. Hal ini memungkinkan perubahan padangan dan respon masyarakat terhadap pembangunan tersebut.

Penelitian ini bertujuan untuk mengkaji bagaimana masyarakat sekitar merespon adanya pembangunan tersebut melalui analisis sentimen terhadap beberapa komentar instagram dengan metode *Support Vektor Machine.* Penelitian ini bersifat studi eksploratif awal, dikarenakan keterbatasan peneliti dalam mengakses data serta keterbatasan dalam menyeimbangkan datanya. Meskipun demikian, peneliitian ini diharapkan bisa memberikan gambaran awal mengenai persepsi masyarakat terhadap pembangunan Kawasan Industri Terpadu Batang serta membukaruang kajian lebih lanjut dengan data yang luas dan hasil yang seimbang. Analisis sentimen digunakan untuk memahami opini publik berbasis teks. Studi sebelumnya menunjukkan dampak sosial-ekonomi pembangunan industri dan efektivitas algoritma SVM dalam klasifikasi sentimen. TF-IDF juga umum digunakan sebagai metode ekstraksi fitur dalam analisis teks. Pendekatan ini relevan dalam menilai opini masyarakat secara real-time melalui media sosial.

2 METODE

Pengumpulan data dalam penelitian ini dilakukan melalui studi literatur dan analisis dokumen dengan fokus pada komentar manual dari Instagram terkait pembangunan Kawasan Industri Terpadu (KIT) Batang dengan postingan dari tahun 2022 sampai 2025. Data tersebut dihimpun untuk menggambarkan dampak sosial, ekonomi, dan lingkungan yang dirasakan masyarakat sekitar, serta menjadi dasar dalam merumuskan rekomendasi yang solutif dan berkelanjutan (Hadi & Sugiarto, 2025). Pene;itian ini Menggunakan metode analisis sentimen pendekatan kualitatif dan kuantitatif untuk mengevaluasi persepsi masyarakat terhadap pembangunan KIT Batang melalui komentar di media sosial. Media sosial dipilih karena mencerminkan opini publik secara real-time dan alami, sehingga dapat mengidentifikasi dukungan maupun penolakan terhadap proyek strategis nasional ini.

Untuk meningkatkan akurasi analisis, digunakan algoritma *Support Vector Machine* (SVM) yang mengklasifikasikan sentimen ke dalam dua kategori: positif dan negatif. Algoritma *Support Vector Machine* (SVM) yaitu algoritma menemukan ruang pemisah pada kumpulan data pada kelas yang tidak sama (Supian et al., 2024). SVM menjadi algoritma yang paling sering digunakan pada awal tahun 90'an hingga awal 2000'an (Agastya, 2018). Hasil analisis ini diharapkan menjadi masukan penting bagi pemangku kebijakan dalam merespons dinamika publik dan mendukung keberhasilan PSN KIT Batang.

2.1 Pengumpulan Data Ulasan Pengguna

Tahap pertama dalam penelitian ini adalah pengumpulan data komentar pengguna Instagram mengenai pembangunan Kawasan Industri Terpadu (KIT) Batang dari platform Instagram**.** Data dikumpulkan secara manual dengan mencari komentar yang menggunakan hashtag relevan seperti *#grandbtangcity, #kitbatang, #kitb*. Komentar dikumpulkan dari postingan 15 Oktober sampai 16 April 2025.

2.2 Preprocecing Data

Setelah data ulasan pengguna berhasil dikumpulkan, langkah selanjutnya yang dilakukan adalah tahap *preprocessing data* atau prapemrosesan data. Tahapan ini sangat penting dalam penelitian analisis sentimen karena bertujuan untuk membersihkan data untuk menghapus karakter khusus, normalisasi teks untuk mengubah teks menjadi format standar, penghapusan kata-kata yang tidak relevan dan menyiapkan data agar dapat dianalisis secara lebih akurat dan sistematis(Yanuar, 2024). Pada penelitian ini, preprocessing dilakukan melalui beberapa tahapan yaitu Pertama, menyiapkan berbagai pustaka (*library*) *Python* yang akan digunakan dalam proses analisis. Kedua, membaca data. Ketiga, dilakukan proses pelabelan (labeling) terhadap komentar (*positif* atau *negatif*) berdasarkan isi atau nada komentar tersebut. Keempat adalah menghitung jumlah data yang mengandung nilai kosong (NULL), yang dapat memengaruhi hasil analisis jika tidak ditangani. Dengan langkah-langkah ini, data yang awalnya tidak terstruktur dapat diproses menjadi lebih bersih, seragam, dan siap digunakan dalam tahap analisis selanjutnya

2.2.1 Menyiapkan *Libary* yang digunakan

Langkah pertama dalam mendeskripsikan tools yang digunakan dalam penelitian ini adalah mengidentifikasi tools visualisasi data yang dibutuhkan, seperti *Pandas* untuk manipulasi data, *Matplotlib* dan *WordCloud* untuk visualisasi data, dan *Scikit-learn* untuk pemodelan dan evaluasi. Setelah itu, package manager, seperti pip, digunakan untuk menginstall pustaka-pustaka yang telah disebutkan di atas. Beberapa contoh yang digunakan antara lain import pandas sebagai PDB, *import matplotlib.pyplot* sebagai PLT, *import WordCloud*, dan berbagai *tools Scikit-learn* seperti *train_test_split, TfidfVectorizer, LinearSVC*, dan metrik evaluasi seperti *accuracy_score* dan *confusion_matrix*. Setelah instalasi, letakkan pustaka dalam kode dan pastikan semuanya bekerja dengan baik, terutama jika Anda menggunakan lingkungan virtual untuk mengatasi masalah kompatibilitas.

2.2.2 Membaca Data

Proses pembacaan data pada penelitian ini dilakukan di *Google Collab* dengan memanfaatkan pustaka Python yang mendukung pengolahan data. File data dalam format CSV diunggah secara manual dengan kode di *Google Colab* menggunakan modul yang tersedia. Setelah berhasil diunggah, file tersebut dibaca dan dimuat ke dalam program untuk selanjutnya diproses pada tahap analisis, seperti pembersihan data dan penerapan metode klasifikasi.

2.2.3 Menghitung jumlah nilai NULL

Menghitung nilai NULL dilakukan sebelum analisis selanjutnya . dalam menghitung nilai NULL, data akan diperiksa untuk mengetahui adanya nilai kosong, dengan menggunakan fungsi *data.isnull( ). Sum ( ),* untuk memastikan kualitas data yang digunakan dalam pemodelan.

2.2.4 Mengubah semua huruf pada data komentar menjadi huruf kecil

Tahap selanjutnya, dilakukan proses normalisasi teks dengan mengubah seluruh huruf pada data komentar menjadi huruf kecil. Langkah ini bertujuan untuk menyamakan format teks agar model tidak membedakan kata yang sama namun ditulis dengan kapitalisasi berbeda, sehingga hasil analisis menjadi lebih konsisten dan akurat.

2.2.5 Menghapus karakter spesial pada data komentar

Proses selanjutnya dalam tahap pra-pemrosesan, dilakukan penghapusan karakter spesial pada data komentar seperti karakter spesial (tanda baca, simbol, dan karakter non-alfabet lainnya dihapus dari data komentar). Proses ini dilakukan guna membersihkan teks dari elemen yang tidak memiliki makna bahasa yang dapat mengganggu kinerja model analisis teks. Dengan menghilangkan karakter-karakter tersebut, data menjadi lebih bersih dan siap untuk dianalisis secara lebih efektif.

2.3 EDA (*Exploratory Data Analysis*)

Dalam proses *Exploratory Data Analysis* (EDA), sejumlah tahapan dilakukan untuk memahami struktur dan pola dalam data sebelum masuk ke tahap analisis lanjutan. Tahap ini bertujuan untuk memahami distribusi data, mendeteksi pola, serta mengidentifikasi anomali seperti entri duplikat atau data yang tidak relevan (Salma et al., 2025). Tahap awal EDA dimulai dengan pelabelan data komentar berdasarkan sentimen menjadi dua kategori utama, yaitu komentar positif dan komentar negatif. Setelah itu, dilakukan visualisasi proporsi masing-masing kategori untuk mengetahui sebaran dan keseimbangan data. Informasi ini penting untuk memastikan bahwa model yang akan dibangun tidak bias terhadap salah satu kelas.

Analisis kemudian dilanjutkan dengan menghitung sepuluh kata yang paling sering muncul dalam komentar positif dan negatif. Hasil perhitungan tersebut divisualisasikan dalam bentuk diagram batang untuk memudahkan interpretasi kata-kata dominan pada masing-masing kategori. Selain itu, dilakukan visualisasi dalam bentuk *word cloud* untuk komentar positif dan negatif secara terpisah.

2.4 Feature Engineering

Proses *feature engineering* meliputi beberapa langkah penting. Salah satu di antaranya adalah pembagian data menjadi dua group: data training dan data testing. Pembagian ini dipakai agar model dapat diuji terhadap data yang tidak pernah dilihat sebelumnya, sehingga dapat mengukur kemampuan model dalam melakukan generalisasi. Selain itu, teknik pembobotan seperti TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk memberikan bobot pada kata pada dokumen (Putri et al., n.d.). Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu yang disebut *term frequency* (TF) dan *inverse frekuensi* dokumen yang mengandung kata yang disebut *inverse document frequency* (IDF). Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata tersebut di dalam dokumen. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut akan rendah pada kumpulan dokume(Sari et al., 2021).

AFTA kemudian menghitung TF-IDF pada data testing. Nilai IDF yang perhitungan menggunakan data training dimasukkan dalam data testing. Langkah ini sangat penting untuk memelihara konsistensi dalam representasi data, oleh karena itu model tida k "mengetahui" informasi dari data tes selama model dilatih.

2.5 Modelling dan Evaluation

Pemodelan ini dilakukan dengan menggunakan algoritma *Support Vector Machine (SVM)* yang diimplementasikan melalui LinearSVC dari library scikit-learn. Model ini dipilih karena kekuatannya dalam memproses data teks dimensi tinggi yang telah diterjemahkan ke dalam representasi angka dengan menggunakan teknik TF-IDF. Model ini juga sangat cocok

Langkah awal dari proses pemodelan adalah pembangunan objek model LinearSVC yang kemudian dilanjutkan dengan pelatihan (*fit*) dengan menggunakan data pelatihan sebelumnya

(*X_train_tfidf dan y_train*). Pada akhir proses pelatihan ini, model diaplikasikan pada proses memprediksi sentimen terhadap data pengujian (*X_test_tfidf*). Prediksi hasil disimpan di *variabel y_pred*. Pengujian performa model dilakukan dengan menghitung nilai akurasi, *recall*, dan presisi menggunakan fungsi *accuracy_score, recall_score*, dan *precision_score*. Selain itu, juga ditampilkan classification report untuk memberikan gambaran yang lebih lengkap mengenai kinerja model dalam memprediksi tiap kelas sentimen. Pengujian ini sangat penting untuk mengetahui sejauh mana model mampu membedakan sentimen positif dan negatif secara akurat.

Analisis evaluasi dilakukan dengan akurasi (*accuracy*) yaitu salah satu metrik yang umum digunakan untuk mengukur sejauh mana model klasifikasi dapat mengklasifikasikan data dengan benar (Sofyan et al., 2024). Matriks akurasi disini menggunakan confusion matrix yang dihasilkan dengan menggunakan fungsi *confusion_matrix()* yang biasanya digunakan untuk mengukur accuracy, recall, precision dan f1-score. Nilai *recall* dan *precision* dihitung untuk mengetahui seberapa baik teknik yang dikembangkan dapat menemukan kasus sentimen di setiap kelasnya. (Putri et al., n.d.)

3 HASIL DAN PEMBAHASAN

Pengumpulan data dalam penelitian ini dilakukan melalui studi literatur dan analisis dokumen dengan fokus pada komentar manual dari Instagram terkait pembangunan Kawasan Industri Terpadu (KIT) Batang. Data tersebut dihimpun untuk menggambarkan dampak sosial, ekonomi, dan lingkungan yang dirasakan masyarakat sekitar, serta menjadi dasar dalam merumuskan rekomendasi yang solutif dan berkelanjutan. Menggunakan metode analisis sentimen untuk mengevaluasi persepsi masyarakat terhadap pembangunan KIT Batang melalui komentar di media sosial. Media sosial dipilih karena mencerminkan opini publik secara real-time dan alami, sehingga dapat mengidentifikasi dukungan maupun penolakan terhadap proyek strategis nasional ini. Untuk meningkatkan akurasi analisis, digunakan algoritma *Support Vector Machine* (SVM) yang mengklasifikasikan sentimen ke dalam dua kategori: positif dan negatif. SVM mampu menangani data dalam jumlah besar dan memahami kompleksitas bahasa masyarakat di media sosial. Hasil analisis ini diharapkan menjadi masukan penting bagi pemangku kebijakan dalam merespons dinamika publik dan mendukung keberhasilan PSN KIT Batang.

3.1 Pengumpulan Data Ulasan Pengguna

Tahap pertama dalam penelitian ini adalah pengumpulan data komentar pengguna Instagram mengenai pembangunan Kawasan Industri Terpadu (KIT) Batang dari platform Instagram**.** Data dikumpulkan secara manual dengan mencari komentar yang menggunakan hashtag relevan seperti *#grandbtangcity, #kitbatang, #kitb*. Komentar yang diambil mencakup teks ulasan, username pengguna, post ID, serta hashtag yang digunakan. Pengumpulan data dilakukan secara manual untuk memastikan bahwa komentar yang diambil relevan dengan topik penelitian dan mencerminkan persepsi langsung masyarakat terhadap pembangunan KIT Batang. Dengan menggunakan metode ini, penelitian dapat meminimalkan bias yang mungkin terjadi jika data dikumpulkan dari sumber yang tidak relevan atau kurang terkait dengan subjek penelitian. Data yang terkumpul kemudian akan dianalisis untuk mengidentifikasi sentimen yang terkandung dalam setiap komentar, yang selanjutnya digunakan untuk mengevaluasi dampak sosial dan ekonomi dari pembangunan kawasan ini.

3.2 Preprocecing Data

Setelah data ulasan pengguna berhasil dikumpulkan, langkah selanjutnya yang dilakukan adalah tahap *preprocessing data* atau prapemrosesan data. Tahapan ini sangat penting dalam penelitian analisis sentimen karena bertujuan untuk membersihkan dan menyiapkan data agar dapat dianalisis secara lebih akurat dan sistematis.

```
import pandas as pd
import string

#visualisasi
import matplotlib.pyplot as plt
from collections import Counter
from wordcloud import WordCloud

#modeling
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.svm import LinearSVC
model = LinearSVC()
from sklearn.metrics import accuracy_score, precision_score, recall_score, classification_report,confusion_matrix
```

Gambar 1. Kode Import Library (Istiyani *et al*., 2025)

Berdasarkan gambar 1, bahasa pemrograman yang dipakai adalah bahasa Python yang berisi *Library* dan pemodelan yang akan digunakan. Pertama terdapat proses import library yang berisi pandas dan string untuk manipulasi dan analisis data, serta library visualisasi seperti matplotlib dan wordcloud untuk menampilkan data grafis. Kemudian ada impor modelling seperti *train_test_split* untuk pembagian data, *TfidfVectorizer* dan *CountVectorizer* untuk ekstraksi fitur dari teks, serta *LinearSVC* sebagai model klasifikasi. Kemudian untuk mengukur performa model menggunakan *accuracy_score*, *precision_score*, *recall_score*, *classification_report, dan confusion_matrix*.

	Username	Komentar	Hastag Postingan
0	zukhr98	Hi min, boleh minta tolong di PIN yaa.. Video ...	#grandbtangcity
1	maghfuruddinahmad85	bangga apanaya kalau ga kerja ya tidak makan	#grandbtangcity
2	infobatang	Keren sih ini, Dermaga pelabuhan sedang dibang...	#grandbtangcity
3	hakimjedot11	Keren sekali UMR,nya	#grandbtangcity
4	nnd_afrz	Semoga UMR nya juga ikut berkembang	#grandbtangcity
...	...	...	...
125	afri.anto.1276487	Di kitb hrus bikin bandara min	#kitb
126	diansr.id	Ngasih loker, tapi giliran daftar ga pernah di...	#kitb
127	panepanp	Gedung elit Lapangan pekerjaan syulit. sudah ...	#kitb
128	bachtiaryudiana	@dians_jr kantor utama KITB aja di jakarta bre...	#kitb
129	diansr.id	mungkin loker yang dibuka selama ini cuma gimmick	#kitb

Gambar 2. Hasil data yang dibaca (Istiyani *et al*., 2025)

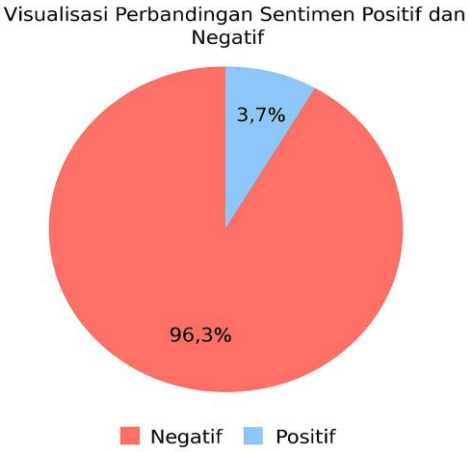
Dari **Gambar 2** menunjukkan data yang diunggah kedalam program tersebut sudah berhasil terbaca. Hasil pembacaan data menunjukkan adanya tiga kolom utama, yaitu username**,** komentar**,** dan hashtag postingan**.** Kolom komentar berisi opini pengguna yang menjadi objek utama analisis sentimen. Sebagian besar komentar, terutama pada hashtag *#kitb*, bernada negatif dan berisi kritik terhadap sistem ketenagakerjaan atau pembangunan. Sementara itu, hashtag *#grandbtangcity* cenderung menampilkan komentar yang lebih positif atau harapan. Informasi ini menjadi dasar penting dalam memahami persepsi publik sebelum dilakukan analisis sentimen lebih lanjut.

Setelah pembacaan data secara manual dilakukan, selanjutnya dilakukan pembersihan data yang meliputi penghitungan nilai kosong (*null*) untuk memastikan tidak ada data yang hilang

pada kolom komentar (Julianto, 2022). Setelah itu, seluruh teks komentar diubah menjadi huruf kecil (*lowercase*) guna menyamakan format kata dan menghindari duplikasi akibat perbedaan kapitalisasi dan menghapus karakter spesial seperti tanda baca dan simbol, agar teks menjadi bersih dan tidak mengandung elemen yang dapat mengganggu proses tokenisasi pada tahap analisis selanjutnya.

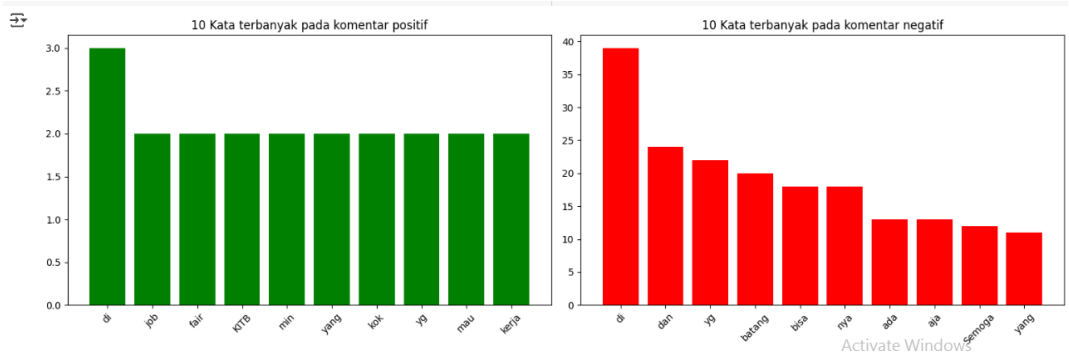
3.3 EDA (*Exploratory Data Analysis*)

Tahap selanjutnya EDA (*Exploratory Data Analysis*), yang dilakukan dengan cara mengeksplorasi data melalui visualisasi seperti visualisasi diagram dan word cloud. EDA juga membantu mengamati proporsi komentar positif dan negatif, kata-kata yang sering muncul, serta konteks umum opini yang terdapat dalam data teks.



Gambar 3. Hasil Visualisasi komentar melalui diagram (Istiyani *et al*., 2025)

Pada Gambar 3, menunjukkan persentase sentimen positif dan negatif yang berhasil diklasifikasikan oleh program dari data yang dianalisis dan divisualisasikan dalam bentuk diagram lingkaran. Diagram tersebut memperlihatkan bahwa 96,3% komentar bersentimen negatif sedangkan 3,7% komentar bersentimen positif. Hal ini menunjukkan bahwa sebagian besar respons masyarakat bersifat negatif, dengan beberapa komentar tentang akses kerja yang sulit, ketidakpuasan atau kehawatiran masyarakat akan dampak pembangunan KIT Batang yang berkelanjutan. Meskipun respon negatif mendominasi masih ada rspon positif yang mendukung pembangunan tersebut.



Gambar 4. Hasil Visualisasi kata terbanyak dalamkomentar positif dan negatif (Istiyani *et al*., 2025)

Pada Gambar 4**,** menunjukkan diagram yang menggambarkan 10 kata yang paling sering muncul dalam komentar positif dan negatif. Pada diagram sebelah kiri bahwa kata yang paling banyak digunakan dalam komentar positif diantaranya "di", "*job*", "*fair*", "*fitur*", "*min*", dan

"yang". Kata-kata ini cenderung bersifat netral hingga positif, menunjukkan kemungkinan adanya apresiasi terhadap suatu argumen. Sementara itu, diagram di sebelah kanan menunjukkan kata-kata yang paling sering muncul dalam komentar negatif, di mana kata "di" dan "dari" mendominasi, diikuti oleh "batang", "tidak", "sama", dan "nya". Kata-kata ini menunjukkan indikasi adanya ketidakpuasan atau kritik terhadap suatu hal, seperti pembangunan di kawasan Industi Terpadu Batang.

Gambar 5. Hasil Word Cloud komentar positif dan negatif (Istiyani *et al*., 2025)

Gambar 5, menunjukkan Hasil *word cloud* yang mempresentasikan hasil dari kata-kata terbanyak yang muncul di beberapa komentar positif dan negatif. *Word cloud* positif menunjukkan harapan dan antusiasme masyarakat terhadap KIT Batang, terutama dalam konteks peluang kerja. Sebaliknya, *word cloud* negatif mencerminkan kekhawatiran, ketidakpuasan, atau kritik, terutama terkait ketidakadilan dalam distribusi lapangan kerja, isu UMR, dan dampak terhadap masyarakat lokal. Perbedaan nuansa kata dalam kedua *word cloud* ini sangat membantu dalam memahami persepsi publik terhadap proyek atau kebijakan yang sedang berjalan.

3.4 Feature Engineering

Pada tahap selanjutnya dilakukan proses pembobotan teks menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF)**,** untuk menilai seberapa penting sebuah kata dalam satu dokumen terhadap seluruh kumpulan dokumen (Albin Pranata et al., 2024). Teknik ini menggabungkan dua komponen utama, yaitu TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*)**.** Dalam metode TF-IDF, nilai TF dan IDF dikalikan bersama-sama untuk menghasilkan bobot kata (*term weight*) untuk setiap kata dalam sebuah dokumen.

	12	180	1jt	1th	30	3000	300t	350	600	70	...	yaa	\
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.234322	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	
..	...	...	...	...	...	...	...	...	...	...	...	...	
127	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	
128	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	
129	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	
130	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	
131	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	

	yakerana	yang	yang150	ybs	yg	yi	yihong	ympe	zukhr98
0	0.0	0.095582	0.0	0.000000	0.0	0.0	0.0	0.0	0.142238
1	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000
2	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000
3	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000
4	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000
..	...	...	...	...	...	...	...	...	...
127	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000
128	0.0	0.000000	0.0	0.205558	0.0	0.0	0.0	0.0	0.000000
129	0.0	0.274810	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000
130	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000
131	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000

[132 rows x 824 columns]

Gambar 6. Hasil Pembobotan dengan TF-IDF (Istiyani *et al*., 2025)

Berdasarkan pada Gambar 6, metode pembobotan menggunakan TF-IDF menghasilkan matriks berdimensi 132 × 824, yang artinya terdapat 132 baris data dan 824 fitur kata unik.

Setiap elemen dalam matriks menunjukkan bobot TF-IDF dari sebuah kata terhadap dokumen tertentu. Sebagian besar nilai dalam matriks bernilai nol, yang mencerminkan karakteristik kata hanya muncul di beberapa dokumen. TF-IDF tinggi menunjukkan bahwa kata itu relatif sering muncul dalam dokumen tersebut melainkan jarang muncul dalam dokumen lain, jadi dinilai penting. Sebaliknya dengan itu, hasil rendah (0) akan menunjukkan bahwa kata tersebut tidak relevan atau umum dalam seluruh korpus. Setiap nilai pada matriks merupakan hasil kalkulasi TF-IDF dari satu kata terhadap satu dokumen. Misalnya:

1. Pada dokumen ke-0, kata "yang" memiliki bobot TF-IDF sebesar 0.95582. Nilai ini menunjukkan bahwa kata "yang" sangat sering muncul dalam dokumen tersebut dan cukup jarang muncul di dokumen lainnya
2. Kata "yaa" memiliki bobot 0.234322, menandakan kata tersebut penting tetapi tidak sebesar kata "yang".
3. Sebagian besar nilai lainnya adalah 0.0, yang berarti kata tersebut tidak muncul dalam dokumen tersebut.

3.5 *Modelling* dan *Evaluation*

```
Akurasi model =  0.9629629629629629
Recall model =  0.9629629629629629
Precision model =  0.9272976680384087
Akurasi model =  0.9629629629629629
Recall model =  0.9629629629629629
Precision model =  0.9272976680384087

Classification Report:
              precision    recall  f1-score   support

    negative       0.96      1.00      0.98        26
    positive       0.00      0.00      0.00         1

    accuracy                           0.96        27
   macro avg       0.48      0.50      0.49        27
weighted avg       0.93      0.96      0.94        27
```

Gambar 7. Hasil *Modelling* dan *Evaluation* (Istiyani *et al*., 2025)

Berdasarkan gambar 7, hasil pemodelan yang dilakukan menunjukkan akurasi 96%, dengan nilai *recall* 96% dan *prescision* 92%. Akan tetapi model ini dipengaruhi oleh ketidakseimbangan data yang digunakan, dimana data negatif lebih besar dibandingkan data positif, hal ini bisa menyebabkan model memprediksi data sebagai kelas negatif. Dari classification report, diketahui bahwa dalam kelas negatif model mampu memberikan hasil yang sangat baik. Hal ini menunjukan data yang bersentimen negatif berhasil diprediksi dengan benar oleh model. Namun, model gagal dalam mengidentifikasi data dengan benar pada kelas positif, yang ditujukan dari nilai *precision*, *recall*, dan *f1-score* yang semua hasilnya 0,00. Hasil tersebut mengidentifikasi bahwa akurasi yang tinggi tidak selalu menjadi faktor keberhasilan model dalam memprediksi, terutama dalam data yang tidak seimbang. Hasil ini perlu dilakukan evaluasi ulang terhadap metode dan analisis yang digunakan serta perbaikan data seimbang agar data bisa mendeteksi keberadaan kedua kategori sentimen secara adil.

4 KESIMPULAN

Penelitian ini menunjukkan bahwa komentar masyarkat bersentimen negatif terhadap pembangunan Pembangunan KIT Batang, dengan persentase 96,3%, sementara persentasekomentar positif hanya 3,7%. Analisis ini dilakukan menggunakan algoritma *Support Sistem Machien (SVM)* yang menghasilkan akurasi tinggi yaitu sebesar 96%. Meskipun akurasi model tinggi penelitian ini memiliki keterbatasan dalam mengeseimbangkan data, sehingga

model hanya bisa mendeteksi komentar negatif tetapi gagal dalam mendeteksi komentar positif. Temuan inibertujuan untuk studi literatur awal untuk mengetahui sejauh mana opini masyarakat terhadapa pembangunan KIT Batang tersebut. Penelitian ini bisa memberikan gambaran bahwa pembangunan masih menimbulkan dampak negatif walaupun dikenyataannya masih ada komentar positif yang tidak terbaca oleh model. Untuk peneliti selanjutnya, disaranakan bisa memperluas data dan keseimbangan data menggunakan teknik optimasi paramater agar temuan lebih akurat dan bisa dijadikan penyusunan kebijakan yang lebih adil dan merata (Darmawan et al., 2023). Penelitian memiliki kontribusi dalam literatur analisis sentimen pembangunan melalui studi kasus di KIT Batang dengan machine learning menggunakan algoritma *Support Vektor Machine*, dimana penelitian ini mampu menangkap beberapa opini masyarakat dari komentar instagram secara *real-time*.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih Kepada pihak Prodi Sains Data dan Fakultas Ekonomi dan Bisnis Islam UIN K.H Abdurrahman Wahid Pekalongan khususnya kepada Ibu Umi Mahmudah M.Sc., Ph.D atas bimbingan dan dukungan yang diberikan dalamproses penyusunan penelitian ini. Terima kasih juga kepada teman-teman dan seluruh pihak yang sudah membantu proses penelitian baik secara langsung maupun tidak langsung.

DAFTAR PUSTAKA

Agastya, I. M. A. (2018). Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film. *Jurnal Tekno Kompak*, *12*(1), 18. https://doi.org/10.33365/jtk.v12i1.70

Albin Pranata, R., Azmi Verdikha, N., Muhammdiyah Kalimantan Timur, U., & Ir Juanda, J. H. (2024). *Metode Pembobotan Tf-Idf Untuk Klasifikasi Teks Quick Count Pemilihan Wakil Presiden Indonesia 2024 Pada X Twitter Dengan Metode Svm*. *18*(2), 126. https://doi.org/10.47111/JTIAvailableonlineathttps://e-journal.upr.ac.id/index.php/JTI

Bandorski, D., Kurniawan, N., Baltes, P., Hoeltgen, R., Hecker, M., Stunder, D., & Keuchel, M. (2016). Contraindications for video capsule endoscopy. *World Journal of Gastroenterology*, *22*(45), 9898–9908. https://doi.org/10.3748/wjg.v22.i45.9898

Darmawan, I. P. D. W., Pradnyana, G. A., & Pascima, I. B. N. (2023). Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram. *SINTECH (Science and Information Technology) Journal*, *6*(1), 58–67. https://doi.org/10.31598/sintechjournal.v6i1.1245

Hadi, N., & Sugiarto, D. (2025). *Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM , Logistic Regression dan Naïve Bayes*. *10*(1), 37–49. https://doi.org/10.30591/jpit.v10i1.7106

Industri, K., Kit, T., Nugraha, S. B., Aji, A., & Fauzia, H. (2023). *(Study of the Readiness of Batang Regency Government in Facing the National Strategic Project (Psn)*. 33–40.

Julianto, I. T. (2022). Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut. *Jurnal Algoritma*, *19*(1), 449–456. https://doi.org/10.33364/algoritma/v.19-1.1112

Kusumawati, M. P. (2023). Tinjauan Yuridis Permasalahan dan Tantangan Pengadaan Tanah dalam Kerangka Proyek Strategis Nasional (PSN). *Problematika Kemudahan Proyek Strategi Nasional: Konflik Norma Dan Tantangan Kesejahteraan*, *1*(3), 102–110. https://journal.uii.ac.id/psha/article/view/30962

Meylinda, F. (2021). a Study of Changes in the Spatial Structure of Bitung Citybased on Indicators of People Movement. *Jurnal Spasial*, *8*(3), 2021.

Putri, A. A., Agustian, S., Abdillah, R., Islam, U., Sultan, N., Kasim, S., & Baru, S. (n.d.). *PENERAPAN METODE LOGISTIC REGRESSION UNTUK*. *7*(1), 95–107.

Salma, T. D., Kurniawan, M. F., Darmawan, R., & Basri, A. (2025). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi) Analisis Sentimen Berbasis Transformer : Persepsi Publik terhadap Nusantara pada Perayaan Kemerdekaan Indonesia yang*. *9*(June), 757–764.

Sari, H., Leonarde Ginting, G., & Zebua, T. (2021). Penerapan Algoritma Text Mining Dan Tf-Idf Untuk Pengelompokan Topik Skripsi. *Terapan Informatika Nusantara*, *2*(7), 414–432. https://ejurnal.seminar-id.com/index.php/tin

Sofyan, A., Fitria, R. I., & Isralestina, F. (2024). Analisis Sentimen Masyarakat dalam Pembangunan City Walk Kota Tegal di Media Sosial Facebook Menggunakan Algoritma Naïve Bayes dan SVM. *Smart Comp: Jurnalnya Orang Pintar Komputer*, *13*(1), 229–234. https://doi.org/10.30591/smartcomp.v13i1.5785

Supian, A., Tri Revaldo, B., Marhadi, N., Efrizoni, L., & Rahmaddeni, R. (2024). Perbandingan Kinerja Naïve Bayes Dan Svm Pada Analisis Sentimen Twitter Ibukota Nusantara. *Jurnal Ilmiah Informatika*, *12*(01), 15–21. https://doi.org/10.33884/jif.v12i01.8721

Yanuar, R. A. A. (2024). *Jurnal Teknik Informatika, Vol. 16, No. 2, April 2024*. *16*(2), 1–7. https://ejurnal.ulbi.ac.id/index.php/informatika/article/view/3533