

Analisis Perbandingan Kinerja TextBlob dan CLIFF NLP pada Klasifikasi Sentimen Ulasan Pelanggan Pengrajin Jepara

Hendra Achmadi, Sylvia Samuel, Zoel Hutabarat, Yorkie Radnan Kristiono

Abstrak

Analisis sentimen merupakan salah satu teknik dalam pemrosesan bahasa alami (Natural Language Processing/NLP) yang digunakan untuk mengidentifikasi opini dan emosi pengguna terhadap suatu produk atau layanan berdasarkan data teks. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan konsumen pengrajin mebel di Jepara yang diperoleh dari Google Maps Review, serta membandingkan kinerja dua pendekatan yang berbeda, yaitu metode berbasis transformer dan metode berbasis leksikon. Metode yang digunakan dalam penelitian ini adalah Text Mining-based Sentiment Analysis Method yang meliputi tahapan pengumpulan data, preprocessing teks, ekstraksi fitur, klasifikasi sentimen, dan evaluasi hasil. Pendekatan transformer diimplementasikan menggunakan model `cardiffnlp/twitter-xlm-roberta-base-sentiment` yang mampu memahami konteks kalimat melalui representasi embedding dan mekanisme self-attention, sedangkan pendekatan leksikon menggunakan TextBlob yang mengklasifikasikan sentimen berdasarkan kamus polaritas kata. Hasil analisis terhadap data ulasan menunjukkan bahwa distribusi sentimen didominasi oleh sentimen negatif dengan 63 ulasan (54%), diikuti oleh sentimen positif sebanyak 38 ulasan (33%), dan sentimen netral sebanyak 16 ulasan (13%). Temuan ini menunjukkan bahwa sebagian konsumen masih memiliki pengalaman yang kurang memuaskan terhadap layanan atau produk pengrajin Jepara. Secara metodologis, model berbasis transformer menunjukkan kemampuan yang lebih baik dalam memahami konteks bahasa dibandingkan metode leksikon yang cenderung bergantung pada polaritas kata secara individual. Penelitian ini diharapkan dapat memberikan kontribusi dalam memahami persepsi konsumen terhadap industri kerajinan lokal serta memberikan referensi metodologis dalam penerapan analisis sentimen pada data ulasan berbahasa Indonesia.

Kata kunci: sentiment analysis, text mining, transformer model, XLM-RoBERTa, TextBlob, ulasan konsumen.

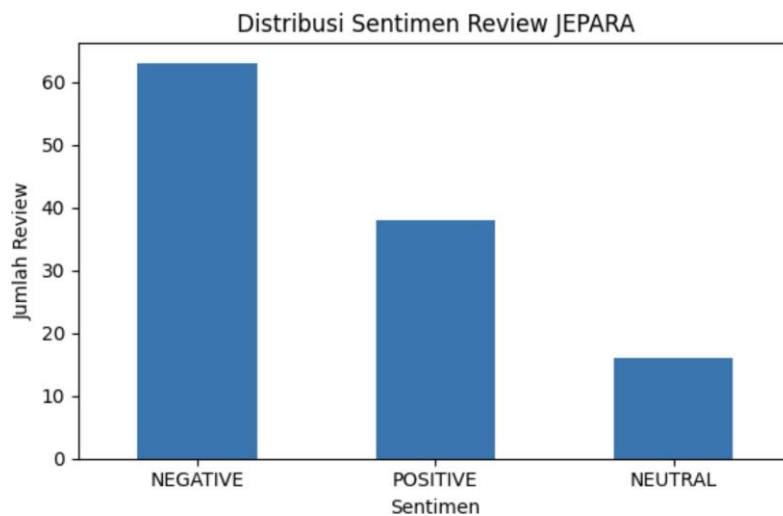
Pendahuluan

Analisis sentimen merupakan salah satu cabang penting dalam pemrosesan bahasa alami (Natural Language Processing/NLP) yang bertujuan untuk mengidentifikasi opini, emosi, dan sikap pengguna terhadap suatu objek atau layanan yang diekspresikan melalui teks. Perkembangan analisis sentimen mengalami transformasi signifikan, dari pendekatan berbasis leksikon menuju metode berbasis pembelajaran mesin dan deep learning (Liu, 2012; Medhat et al., 2014). Pendekatan awal yang mengandalkan kamus sentimen relatif sederhana dan efisien, namun memiliki keterbatasan dalam memahami konteks kalimat, negasi, serta ambiguitas bahasa. Seiring berkembangnya arsitektur neural, model berbasis transformer seperti BERT dan turunannya menunjukkan peningkatan kinerja yang signifikan dalam berbagai tugas NLP karena mampu memahami representasi kontekstual kata secara lebih baik (Devlin et al., 2019; Young et al., 2018).

Dalam konteks industri kreatif di Indonesia, khususnya pada sektor kerajinan mebel di Jepara, opini konsumen yang disampaikan melalui platform digital seperti Google Review menjadi sumber informasi penting untuk memahami persepsi pasar terhadap kualitas produk dan layanan pengrajin. Berdasarkan data review yang dikumpulkan melalui Google Form dari konsumen

pengrajin di Jepara, terlihat bahwa distribusi sentimen menunjukkan dominasi sentimen negatif dibandingkan sentimen positif dan netral. Fenomena ini mengindikasikan adanya ketidakpuasan konsumen terhadap beberapa aspek, seperti kualitas layanan, komunikasi, maupun proses transaksi. Analisis terhadap opini konsumen tersebut menjadi penting untuk memberikan gambaran yang lebih objektif mengenai persepsi pelanggan terhadap industri kerajinan Jepara.

Meskipun berbagai metode analisis sentimen telah banyak dikembangkan, kajian komparatif yang secara sistematis membandingkan kinerja metode berbasis leksikon dengan model transformer pada data ulasan berbahasa Indonesia, khususnya pada sektor industri kerajinan lokal, masih relatif terbatas. Oleh karena itu, penelitian ini bertujuan untuk melakukan perbandingan kinerja antara metode analisis sentimen berbasis leksikon dan model transformer dalam mengklasifikasikan sentimen pada ulasan konsumen pengrajin di Jepara, sehingga dapat memberikan pemahaman yang lebih komprehensif mengenai efektivitas kedua pendekatan tersebut.



Gambar 1. Sentiment Analysis Pengrajin Jepara (Source Google Maps Review)

Berdasarkan hasil pengolahan data ulasan konsumen terhadap pengrajin di Jepara, diperoleh distribusi sentimen yang menunjukkan kecenderungan dominasi sentimen negatif. Dari total ulasan yang dianalisis, sebanyak 63 ulasan diklasifikasikan sebagai sentimen negatif, 38 ulasan sebagai sentimen positif, dan 16 ulasan sebagai sentimen netral. Hasil ini menunjukkan bahwa proporsi sentimen negatif mencapai sekitar 54% dari keseluruhan data, sedangkan sentimen positif sekitar 33% dan sentimen netral sekitar 13%. Dominasi sentimen negatif tersebut mengindikasikan bahwa sebagian konsumen masih memiliki pengalaman yang kurang memuaskan terhadap layanan atau produk yang diberikan oleh pengrajin. Di sisi lain, keberadaan sentimen positif menunjukkan bahwa terdapat pula konsumen yang merasa puas terhadap kualitas produk atau pelayanan yang diterima. Analisis distribusi sentimen ini menjadi dasar

penting untuk memahami persepsi konsumen secara lebih mendalam serta menjadi landasan dalam mengevaluasi kinerja metode analisis sentimen yang digunakan dalam penelitian ini.

Landasan Teori

Penelitian terkini pada abstractive summarization menekankan masalah *factuality* atau *hallucination* pada model generatif berbasis transformer. Model seperti BART sering menghasilkan ringkasan yang secara linguistik lancar namun mengandung informasi yang tidak sesuai dengan dokumen sumber. Untuk mengatasi permasalahan ini, CLIFF (*Contrastive Learning for Improving Faithfulness and Factuality*) diperkenalkan oleh Cao dan Wang (2021) dengan memformulasikan pendekatan *contrastive learning* guna membedakan ringkasan faktual (positif) dan ringkasan yang mengandung kesalahan (negatif). Pendekatan ini bertujuan meningkatkan kesesuaian fakta antara ringkasan dan dokumen sumber. Eksperimen pada dataset Sum dan CNN/DailyMail menunjukkan bahwa CLIFF mampu meningkatkan konsistensi faktual (Cao & Wang, 2021).

Metode CLIFFNLP

Secara algoritmik, CLIFF mengintegrasikan *contrastive loss* pada level sekuens selain maximum likelihood estimation (MLE) standar. Model dilatih untuk memaksimalkan jarak representasi antara ringkasan referensi dan ringkasan negatif yang mengandung kesalahan faktual. Pendekatan ini merupakan bagian dari tren yang lebih luas dalam NLP yang memanfaatkan *contrastive objectives* untuk meningkatkan kualitas representasi dan ketahanan model (Ströhle et al., 2024). Perbandingan Tujuan dan Ruang Lingkup Berbeda dengan CLIFF yang bersifat *supervised* dan dirancang untuk tugas generatif berisiko tinggi terhadap kesalahan informasi (misalnya ringkasan berita atau medis), TextBlob merupakan toolkit NLP ringan berbasis Python yang menyediakan analisis sentimen berbasis *leksikon*, *part-of-speech tagging*, dan berbagai utilitas pemrosesan teks lainnya. TextBlob mengandalkan kamus polaritas dan *heuristik* linguistik sehingga mudah diimplementasikan dan efisien secara komputasi. Namun, pendekatan ini tidak memanfaatkan pembelajaran representasi berbasis data besar sehingga rentan terhadap domain shift, negasi kompleks, ironi, dan variasi bahasa tidak baku (Ashbaugh et al., 2024). Studi komparatif menunjukkan bahwa metode leksikon seperti TextBlob masih relevan sebagai baseline, tetapi umumnya memiliki akurasi lebih rendah dibandingkan model berbasis transformer pada dataset yang kompleks atau noisy (Ashbaugh et al., 2024).

Evaluasi pada metode berbasis *contrastive learning* seperti CLIFF tidak hanya mengandalkan metrik, tetapi juga menggunakan metrik berbasis pertanyaan-jawaban (QA-based factuality) serta evaluasi manusia untuk mengukur konsistensi fakta (Cao & Wang, 2021). Literatur survei terbaru menunjukkan bahwa penerapan *contrastive objectives* secara konsisten meningkatkan *factuality* tanpa mengorbankan kelancaran teks, khususnya ketika strategi *negative sampling* dirancang dengan baik (Ströhle et al., 2024).

Metode TextBlob

Sebaliknya, evaluasi TextBlob dalam berbagai studi menunjukkan tingkat akurasi sedang (sekitar 0.60–0.70 pada beberapa dataset ulasan), dengan kelemahan utama pada recall untuk sentimen tertentu dan sensitivitas terhadap konteks linguistik yang kompleks (Ashbaugh et al., 2024).

Kekuatan dan Keterbatasan CLIFF unggul dalam konteks yang memerlukan akurasi faktual tinggi dan tersedia data pelatihan berlabel, namun memerlukan sumber daya komputasi besar serta desain sampel negatif yang cermat (Cao & Wang, 2021). Sebaliknya, TextBlob menawarkan kemudahan penggunaan, kecepatan, dan interpretability, tetapi memiliki keterbatasan dalam menangani variasi bahasa kompleks dan konteks domain-spesifik (Ashbaugh et al., 2024).

Dengan demikian, pemilihan metode harus disesuaikan dengan kebutuhan aplikasi. Untuk sistem yang menuntut keakuratan informasi tinggi, pendekatan contrastive seperti CLIFF lebih direkomendasikan. Namun, untuk prototyping cepat atau analisis awal pada data yang relatif bersih, TextBlob tetap layak digunakan sebagai baseline.

Methodological Sentiment Analysis

Text Mining-based Sentiment Analysis Method merupakan teknik dalam text mining yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini atau emosi dalam teks menjadi kategori seperti positif, negatif, atau netral. Tahapan metodologi sentiment analysis umumnya dimulai dari pengumpulan data, misalnya dari media sosial, ulasan pelanggan, atau forum daring. Tahap berikutnya adalah preprocessing, yang meliputi pembersihan teks (cleaning), tokenisasi, normalisasi, stopword removal, dan stemming untuk mengurangi noise dan meningkatkan kualitas fitur (Liu, 2012). Selanjutnya dilakukan ekstraksi fitur, seperti word embeddings. Pada tahap pemodelan, algoritma machine learning seperti Naïve Bayes, Support Vector Machine (SVM), maupun pendekatan deep learning seperti LSTM atau transformer digunakan untuk melakukan klasifikasi sentimen (Medhat, Hassan, & Korashy, 2014). Setelah model dibangun, dilakukan evaluasi model menggunakan metrik seperti accuracy, precision, recall, dan F1-score untuk mengukur performa klasifikasi. Tahap akhir adalah interpretasi dan analisis hasil, yang bertujuan untuk menarik insight strategis bagi pengambilan keputusan. Pemilihan metode sangat bergantung pada karakteristik data, bahasa, serta kebutuhan analisis yang dilakukan.

Metode yang digunakan dalam penelitian ini adalah Text Mining-based Sentiment Analysis Method, yang memanfaatkan teknik pemrosesan bahasa alami (Natural Language Processing/NLP) untuk mengekstraksi dan mengklasifikasikan opini dari data teks. Text mining merupakan pendekatan yang banyak digunakan untuk mengidentifikasi pola dan informasi yang tersembunyi dalam data tidak terstruktur seperti ulasan pengguna atau komentar konsumen (Liu, 2012). Dalam konteks analisis sentimen, proses ini biasanya melibatkan beberapa tahapan utama yaitu pengumpulan data, pembersihan data (data cleaning), preprocessing teks, representasi fitur, dan klasifikasi sentimen (Medhat et al., 2014). Tahap preprocessing dilakukan untuk meningkatkan kualitas data melalui proses seperti case folding, tokenization, stopword removal, dan stemming sehingga teks lebih mudah diproses oleh algoritma analisis sentimen (Manning et

al., 2008). Selanjutnya, teks direpresentasikan ke dalam bentuk numerik menggunakan metode seperti Bag-of-Words atau TF-IDF sebelum dilakukan proses klasifikasi sentimen ke dalam kategori positif, negatif, atau netral. Perkembangan terbaru dalam analisis sentimen juga memanfaatkan model berbasis deep learning dan arsitektur transformer yang mampu memahami konteks bahasa secara lebih baik dibandingkan metode tradisional (Devlin et al., 2019). Dengan demikian, pendekatan text mining berbasis NLP menjadi metode yang efektif untuk menganalisis persepsi konsumen terhadap produk atau layanan yang diekspresikan melalui ulasan teks (Cambria et al., 2017).

Hasil dan Pembahasan sebagai berikut:

1. Pengambilan data dari Google Maps Review

- ✔ Pengrajin ukir Jepara -> ChIJOR_en6kfcS4RVJUGs8y2uRc
- ✔ Pengrajin mebel Jepara -> ChIJq6qq6uMecS4Rd8N942TCQg8
- ✔ Pengrajin kayu Jepara -> ChIJ1VowX4b3aS4R0QajUJJBbbu
- ✔ Toko mebel Jepara -> ChIJq6qq6uMecS4Rd8N942TCQg8
- ✔ Furniture factory Jepara -> ChIJq-US_4EfcS4RFneqnTzy30o
- ✔ Showroom mebel Jepara -> ChIJq6qq6uMecS4Rd8N942TCQg8
- ✔ Pengrajin kursi Jepara -> ChIJq6qq6uMecS4Rd8N942TCQg8
- ✔ Pengrajin almari Jepara -> ChIJyXJiLy0fcS4RjJivrERYctk
- ✔ Pengrajin dipan Jepara -> ChIJWam4T5nhcC4R08dL7dDTOrs
- ✔ Workshop furniture Jepara -> ChIJTzASMywfcS4RXTi9KuK-fj8
- ✔ Pengrajin furniture jati Jepara -> ChIJq6qq6uMecS4Rd8N942TCQg8
- ✔ Pengrajin mebel jati ukir Jepara -> ChIJq6qq6uMecS4Rd8N942TCQg8
- ✔ Pabrik furniture Jepara -> ChIJf_yxmxAfcS4RBqB40a1gzL8
- ✔ Produsen furniture Jepara -> ChIJ6UV4ZvkecS4ROzuMUaPb1Q4
- ✔ Supplier furniture Jepara -> ChIJ6UV4ZvkecS4ROzuMUaPb1Q4

2. Mempersiapkan model

```

model_name = "indobenchmark/indobert-base-p1"

print("⌚ Memuat model sentimen IndoBERT, mohon tunggu...")

tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)
classifier = pipeline("sentiment-analysis", model=model, tokenizer=tokenizer)

print("✔ Model siap digunakan")
    
```

	time	author	rating	text	branch
0	2024-06-30	suyati marsudi	5	Preserving one of Indonesia's cultures	Pengrajin ukir Jepara
1	2019-10-21	Unqisa Cap	4	Awesome	Pengrajin ukir Jepara
2	2019-07-23	Finza Jepara	4	NaN	Pengrajin ukir Jepara
3	2018-08-22	Miftah Udin	5	Truths	Pengrajin ukir Jepara
4	2017-12-20	Fenny Sakura	4	Selling various kinds of carvings	Pengrajin ukir Jepara
...
112	2020-12-23	Arini Rusydah	5	Handmade creative product	Sentra furniture Jepara
113	2020-11-13	Jani Ali	5	NaN	Sentra furniture Jepara
114	2019-11-26	Chanel Pelangi Garang	3	reasonable	Sentra furniture Jepara

12/13/2025 Pengrajin furniture jati Jepara Quality is excellent and recommended to who is looking satisfaction.

3. “Quality is excellent and recommended to those who are looking for satisfaction.”
4. Jika dianalisis menggunakan model cardiffnlp/twitter-xlm-roberta-base-sentiment (yang sering disebut CLIFF NLP oleh sebagian pengguna), maka hasilnya kemungkinan besar adalah: POSITIVE

- a. Teks masuk ke model XLM-RoBERTa
Teks di-tokenize kemudian masuk ke embedding dan diproses melalui transformer layers.
- b. Output logits (raw score)
Model tidak langsung memberi probabilitas, tetapi menghasilkan 3 nilai numerik (logits), misalnya:
- c. [-1.2, 0.3, 3.8]
- d. Misalkan dari model CLIFF NLP (XLM-RoBERTa) kita mendapatkan logits (contoh simulasi):
- e. $z = [-1.2, 0.3, 3.8]$
- f. Urutan kelas:
 $z_0 = \text{Negative}$
 $z_1 = \text{Neutral}$
 $z_2 = \text{Positive}$
- g. Rumus Softmax

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^3 e^{z_j}}$$

- h. Hitung nilai eksponensial

$$e^{-1.2} = 0.301 \quad e^{0.3} = 1.350 \quad e^{3.8} = 44.701$$

Hitung total penyebut

$$Total = 0.301 + 1.350 + 44.701 \quad Total = 46.352$$

Hitung probabilitas tiap kelas

Negative

$$(Neg) = \frac{0.301}{46.352} = 0.0065$$

Neutral

$$P(Neu) = \frac{1.350}{46.352} = 0.0291$$

Positive

$$P(Pos) = \frac{44.701}{46.352} = 0.9644$$

Model memberikan probabilitas 96.44% Positive, sehingga kalimat: “Quality is excellent and recommended to who is looking satisfaction.” diklasifikasikan sebagai Strong Positive Sentiment.

a, kedekatan makna dikenali karena vektor angkanya mirip secara matematis. Biasanya diukur menggunakan cosine similarity atau jarak Euclidean.

Bagaimana “mirip” itu dihitung?

Misalnya (ilustrasi sederhana 3 dimensi):

excellent = [0.8, 0.6, 0.1] *great* = [0.75, 0.63, 0.05] *terrible* = [-0.7, -0.6, 0.2]

Kita ukur kemiripan dengan Cosine Similarity:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

Hasilnya kira-kira:

- $\cos(\text{excellent}, \text{great}) \approx 0.98$ (sangat dekat)
- $\cos(\text{excellent}, \text{terrible}) \approx -0.95$ (berlawanan arah)

Artinya:

- Vektor *excellent* dan *great* hampir searah
- Vektor *terrible* arahnya berlawanan

Jika menggunakan metode Lexicon (misalnya TextBlob) TextBlob langsung menghasilkan:

$$Polarity \in [-1, 1]$$

Simulasi hasil:

$Polarity \approx +0.8$ sampai $+1.0$

Karena:

- *excellent* (+1.0)
- *recommended* (+0.7)
- *satisfaction* (+0.8)

Rata-rata skor tinggi (sangat positif)

Hasil Diskusi

Metode CLIFF NLP (cardiffnlp/twitter-xlm-roberta-base-sentiment) dan TextBlob memiliki pendekatan yang berbeda dalam sentiment analysis. CLIFF NLP berbasis model transformer XLM-RoBERTa yang menggunakan contextual embedding. Model ini memahami makna kata berdasarkan konteks kalimat secara keseluruhan. Representasi teks diubah menjadi vektor berdimensi tinggi, diproses melalui self-attention, lalu menghasilkan logit yang

ditransformasikan menjadi probabilitas melalui softmax. Keunggulan pendekatan ini adalah kemampuannya menangkap nuansa bahasa, ironi, serta variasi multibahasa dengan akurasi tinggi, terutama pada data media sosial.

Sebaliknya, TextBlob menggunakan pendekatan lexicon-based dengan kamus polaritas kata. Skor sentimen dihitung berdasarkan rata-rata polaritas kata dalam kalimat tanpa mempertimbangkan konteks mendalam. Metode ini lebih sederhana, ringan secara komputasi, dan mudah diimplementasikan, namun kurang akurat pada kalimat kompleks, negasi, atau konteks implisit. Dalam penelitian empiris, model transformer seperti XLM-R umumnya menunjukkan performa lebih tinggi dibanding metode berbasis leksikon, khususnya pada dataset besar dan beragam.

Kesimpulan

Secara umum, CLIFF NLP lebih unggul dalam akurasi dan kemampuan memahami konteks karena memanfaatkan deep learning berbasis transformer. Namun, TextBlob tetap relevan untuk analisis cepat dengan kebutuhan komputasi rendah. Pemilihan metode sebaiknya disesuaikan dengan tujuan penelitian, kompleksitas data, serta sumber daya komputasi yang tersedia.

Referensi

- Ashbaugh, L., Rane, N., & Kumar, A. (2024). A comparative study of sentiment analysis approaches on customer review data. *Computers*, 13(12), 340. <https://doi.org/10.3390/computers13120340>
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2017). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15–21.
- Cao, S., & Wang, L. (2021). CLIFF: Contrastive learning for improving faithfulness and factuality in abstractive summarization. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6633–6649. <https://doi.org/10.18653/v1/2021.emnlp-main.532>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Computational Linguistics*, 45(4), 1–15.
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). Indonesian NLP resources and benchmark datasets: A systematic evaluation. *Journal of Natural Language Engineering*, 26(4), 1–25.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>.
- Ströhle, T., Campos, R., & Jatowt, A. (2024). Contrastive text summarization: A survey. *International Journal of Data Science and Analytics*, 17, 1–24. <https://doi.org/10.1007/s41060-023-00434-4>
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3), 55–75.