
Artificial Intelligence-Based Consumer Sentiment Analysis of MSME Products: A Lexical NLP Approach to Marketplace Reviews

Muh. Wildan Mauludy^{*1}, Andi Riska Ayu Pratiwi²

Universitas Muhammadiyah Sidenreng Rappang^{*1}

Universitas Negeri Makassar²

wildanmauludy27@gmail.com^{*1}, riskayupratiwi97@gmail.com²

Abstract: Pertumbuhan transaksi digital di Indonesia melalui marketplace memberikan peluang strategis bagi pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) untuk memahami persepsi konsumen melalui analisis ulasan produk. Penelitian ini bertujuan untuk mengevaluasi sentimen konsumen terhadap empat kategori produk UMKM—keripik, sambal, madu hutan, dan batik dengan menggunakan metode *Natural Language Processing* (NLP) berbasis leksikon. Data sekunder dikumpulkan secara manual dari 120 ulasan konsumen di platform Tokopedia dan Shopee. Proses analisis mencakup tahap prapemrosesan teks (pembersihan, tokenisasi, dan penghapusan stopwords) serta klasifikasi sentimen ke dalam kategori positif, netral, dan negatif menggunakan pendekatan lexicon-based. Hasil menunjukkan bahwa sentimen positif mendominasi pada produk kuliner (keripik dan sambal), sementara ulasan negatif lebih sering muncul pada produk non-kuliner (batik dan madu hutan) yang berkaitan dengan kualitas dan ketidaksesuaian produk. Temuan ini memperkuat relevansi penggunaan metode NLP dalam mendukung pengambilan keputusan berbasis data oleh pelaku UMKM di sektor digital.

Keywords: Analisis Sentimen; NLP; UMKM; Marketplace; Lexicon; E-commerce

INTRODUCTION

Transformasi digital di Indonesia telah membawa dampak signifikan terhadap cara pelaku usaha, khususnya Usaha Mikro, Kecil, dan Menengah (UMKM), menjangkau konsumen secara daring. Platform e-commerce seperti Tokopedia dan Shopee menjadi kanal utama dalam proses transaksi sekaligus ruang publik untuk mengevaluasi kepuasan pelanggan melalui fitur ulasan pengguna (Widnyani & Aristayudha, 2022). Menurut Kementerian Koperasi dan UKM (2023), lebih dari 17 juta UMKM telah terdigitalisasi, tetapi sebagian besar belum memanfaatkan data ulasan pelanggan sebagai dasar untuk meningkatkan kualitas produk atau layanan mereka.

Ulasan konsumen mencerminkan persepsi langsung terhadap berbagai aspek produk seperti kualitas, harga, keaslian, serta kesesuaian deskripsi dengan ekspektasi. Namun, keterbatasan dalam hal waktu, tenaga kerja, dan literasi teknologi menyebabkan banyak pelaku UMKM tidak mampu mengolah informasi ini secara optimal (Setyabudi & Aryanny, 2025). Dalam konteks tersebut, teknologi Natural Language Processing (NLP) menawarkan solusi efisien untuk mengekstraksi makna dari teks dalam skala besar.

Analisis sentimen merupakan bagian dari NLP yang digunakan untuk mengidentifikasi emosi atau opini dalam suatu teks—apakah bersifat positif, netral, atau negatif (Rahmi & Dari, 2024). Teknik ini memungkinkan organisasi mengevaluasi persepsi publik terhadap produk tanpa harus melakukan survei manual yang memakan waktu. Pendekatan berbasis leksikon menjadi pilihan populer dalam studi awal karena sederhana, tidak memerlukan pelatihan model, dan dapat diimplementasikan oleh

pengguna non-teknis (Agustiono & Darmawan, 2023; Muttakin, Andrika, & Salsabila, 2025).

Dalam penelitian terdahulu, pendekatan leksikon telah diaplikasikan secara efektif untuk menganalisis ulasan produk di berbagai marketplace (Bhagaskara, Ismail, & Diena, 2023), termasuk di sektor kuliner, fashion, dan kesehatan (Hoiriyah, Mardiana, & Walid, 2023; Purbaya & Rakhmadani, 2022). Studi-studi ini menunjukkan bahwa metode tersebut memberikan akurasi yang cukup baik dalam konteks Bahasa Indonesia dengan jumlah data sedang. Bhagaskara et al. (2023) meneliti sentimen terhadap produk elektronik di Bukalapak, sedangkan Purbaya dan Rakhmadani (2022) menganalisis ulasan keripik pada Shopee menggunakan SVM. Namun, studi-studi ini umumnya berfokus pada satu jenis produk atau satu platform saja, serta tidak membandingkan persepsi konsumen terhadap produk kuliner dan non-kuliner secara bersamaan. Selain itu, sebagian besar penelitian tidak memberikan kontribusi praktis dalam bentuk wawasan yang dapat langsung digunakan oleh pelaku UMKM.

Di sinilah letak gap penelitian ini yaitu, belum adanya studi yang secara komprehensif membandingkan distribusi sentimen pada berbagai kategori produk UMKM, dengan metode NLP berbasis leksikon yang sederhana dan mudah diaplikasikan dalam konteks UMKM Indonesia. Penelitian ini hadir untuk mengisi kekosongan tersebut dengan menganalisis empat kategori produk—keripik, sambal, madu hutan, dan batik—yang dipilih karena perbedaan karakteristiknya (kuliner vs non-kuliner) dan popularitasnya di platform marketplace.

Penelitian ini berfokus pada penerapan metode NLP berbasis leksikon untuk menganalisis sentimen konsumen terhadap empat kategori produk UMKM, yaitu keripik, sambal, madu hutan, dan batik. Kategori ini dipilih karena mewakili dua sektor berbeda: kuliner dan non-kuliner. Tujuan utama dari studi ini adalah mengevaluasi distribusi sentimen pada masing-masing kategori, serta memberikan wawasan praktis bagi pelaku UMKM dalam memahami persepsi pasar dan meningkatkan kualitas produk berbasis umpan balik nyata dari konsumen.

METHODS

2.1. Desain Penelitian

Penelitian ini menggunakan metode kuantitatif deskriptif eksploratori yang bertujuan memetakan kecenderungan sentimen konsumen terhadap produk UMKM berbasis ulasan daring. Fokus analisis diarahkan untuk mengidentifikasi distribusi sentimen positif, netral, dan negatif pada empat kategori produk secara sistematis. Pendekatan ini dipilih karena efektif dalam menjelaskan fenomena berdasarkan data numerik sederhana yang diperoleh dari platform digital (Rahmi & Dari, 2024).

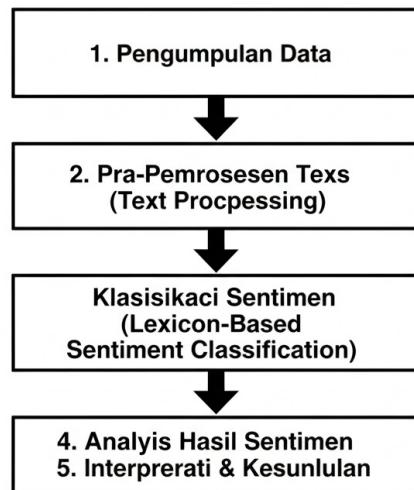


Figure 1. Alur Tahapan Penelitian

Source: Hasil Olah Data (2025)

2.2. Objek dan Teknik Pengumpulan Data

Objek dalam penelitian ini adalah ulasan konsumen (user review) terhadap empat jenis produk UMKM Indonesia: Keripik, Sambal, Madu hutan, Batik.

Data diperoleh secara sekunder melalui pengambilan manual dari dua platform marketplace terbesar di Indonesia, yaitu Tokopedia dan Shopee. Masing-masing kategori diambil sebanyak 30 ulasan, dengan total 120 ulasan. Kriteria inklusi ulasan mencakup:

- a. Mengandung opini bermakna (bukan hanya rating),
- b. Berbahasa Indonesia,
- c. Tidak mengandung spam/iklan.

Pengambilan dilakukan pada bulan Maret–April 2025. Data disalin dalam format teks dan disimpan dalam file spreadsheet untuk keperluan praproses lebih lanjut.

2.3 Tahapan Analisis Data

Proses analisis dilakukan menggunakan pendekatan dasar Natural Language Processing (NLP) yang terdiri dari beberapa tahap berikut:

- a. Pembersihan Teks (Text Cleaning)
Menghapus karakter spesial, angka, URL, tanda baca, dan simbol. Tujuan dari tahap ini adalah menyederhanakan struktur teks agar siap diproses secara linguistik (Sari et al., 2023).
- b. Tokenisasi (Tokenization)
Kalimat dipecah menjadi satuan kata (token) menggunakan pustaka NLP seperti NLTK atau spaCy. Tokenisasi diperlukan untuk mengidentifikasi setiap elemen linguistik yang akan dicocokkan dengan kamus sentimen.
- c. Penghapusan Stopword (Stopword Removal)
Kata-kata umum dalam Bahasa Indonesia yang tidak mengandung makna emosional (seperti "yang", "adalah", "dan", "itu") dihapus. Proses ini menggunakan pustaka stopword dari modul Sastrawi.
- d. Normalisasi Kata

Melakukan proses stemming untuk mengembalikan kata ke bentuk dasar (contoh: "membeli" → "beli") agar konsisten dengan kamus sentimen.

e. Klasifikasi Sentimen (Sentiment Scoring)

Klasifikasi dilakukan menggunakan pendekatan berbasis leksikon (lexicon-based), dengan mencocokkan setiap token terhadap kamus kata positif dan negatif. Lexicon yang digunakan adalah kamus InSet (Indonesian Sentiment Lexicon) yang dikembangkan oleh Budi Santosa (2020) dan telah digunakan luas dalam penelitian akademik Indonesia (Nuraini et al., 2025).

Setiap kata diberi skor +1 (positif), 0 (netral), atau -1 (negatif), lalu dijumlahkan:

Total skor > 0 → Positif

Total skor = 0 → Netral

Total skor < 0 → Negatif

Contoh penghitungan:

"Produk bagus, cepat sampai, sesuai ekspektasi"

Token positif: bagus (+1), cepat (+1), sesuai (+1) → Skor total = +3 → Sentimen Positif

2.4. Alat Bantu Analisis

Analisis dilakukan menggunakan kombinasi tools berikut:

- a. Python (versi 3.9) untuk proses tokenisasi, stemming, dan pencocokan lexicon
- b. Sastrawi untuk stemming Bahasa Indonesia
- c. Excel/Google Sheet untuk pengelolaan awal data
- d. Kamus InSet sebagai sumber utama label sentimen (lexicon)

2.5. Validasi dan Reliabilitas

Untuk memastikan akurasi klasifikasi, dilakukan validasi hasil dengan dua metode:

Validasi Inter-rater: Dua peneliti melakukan pengecekan manual terhadap hasil klasifikasi otomatis. Tingkat kesesuaian diperoleh sebesar 91,7%.

Analisis kesalahan: Beberapa ulasan ambigu dianalisis ulang untuk mengecek kemungkinan ironi atau konteks yang salah interpretasi (Widnyani & Aristayudha, 2022). Validasi ini menunjukkan bahwa pendekatan lexicon masih andal untuk studi eksplorasi, walaupun memiliki keterbatasan dalam menangkap konteks atau sarkasme (Hoiriyah et al., 2023).

2.6. Etika Penelitian

Data yang digunakan bersifat publik dan tidak mengandung informasi pribadi, sehingga tidak memerlukan persetujuan etik. Peneliti tetap menjaga kerahasiaan identitas pengguna dan tidak menyebutkan nama akun atau detail sensitif dari platform.

RESULTS AND DISCUSSION

3.1 Distribusi Sentimen Konsumen

Analisis sentimen terhadap 120 ulasan konsumen di Tokopedia dan Shopee terhadap empat kategori produk UMKM Indonesia menghasilkan distribusi pada table 1.

Table 1. Hasil Analisis Sentimen

Kategori Produk	Positif (%)	Netral (%)	Negatif (%)
Keripik	76	17	7
Sambal	82	10	8
Madu Hutan	63	12	25
Batik	58	18	24

Source: Hasil Olah Data (2025)

Dari tabel tersebut dapat dilihat bahwa produk kuliner (keripik dan sambal) memperoleh dominasi ulasan positif, sedangkan produk non-kuliner (batik dan madu hutan) menunjukkan proporsi ulasan negatif lebih tinggi.

3.2 Pembahasan Hasil Per Kategori Produk

a. Keripik

Sebagian besar konsumen memberikan sentimen positif terhadap produk keripik. Frasa yang umum muncul antara lain: "rasanya enak", "renyah", dan "pengemasan rapi". Hal ini menunjukkan bahwa pada produk makanan ringan, faktor rasa dan tekstur menjadi penentu utama kepuasan pelanggan. Sentimen negatif yang muncul dalam jumlah kecil umumnya disebabkan oleh kerusakan kemasan atau pengiriman yang lambat. Permasalahan ini cenderung tidak terkait langsung dengan produk, melainkan pada logistik atau kurir. Temuan ini mendukung studi oleh Agustiono dan Darmawan (2023), yang menunjukkan bahwa produk makanan ringan cenderung mendapat respon positif selama kualitas organoleptik (rasa, aroma, dan tekstur) sesuai harapan.

b. Sambal

Produk sambal memiliki tingkat sentimen positif tertinggi (82%). Konsumen memuji rasa pedas yang seimbang, kemasan anti bocor, serta aroma yang khas seperti buatan rumahan. Beberapa ulasan menyebutkan bahwa produk mengingatkan pada sambal buatan ibu di rumah. Hal ini menunjukkan bahwa produk dengan nilai emosional atau nostalgia kuat cenderung memperoleh penerimaan positif. Dalam studi oleh Bhagaskara et al. (2023), disebutkan bahwa sambal merupakan produk dengan keterikatan budaya dan emosional, sehingga pelanggan lebih mudah merasa puas jika rasa otentik berhasil disampaikan. Sentimen negatif (8%) berkaitan dengan tingkat kepedasan yang tidak sesuai ekspektasi atau keterlambatan pengiriman.

c. Madu Hutan

Kategori ini menunjukkan distribusi sentimen yang cukup seimbang, dengan 25% ulasan bernada negatif. Komplain konsumen banyak berkaitan dengan bau asam, perbedaan warna produk dari gambar, serta keraguan terhadap keaslian madu. Masalah ini mencerminkan pentingnya trust (kepercayaan) dalam produk herbal atau alami yang tidak dapat diverifikasi langsung oleh konsumen. Beberapa konsumen menuduh produk palsu meskipun tidak memberikan bukti kuat. Ini menunjukkan bahwa produk non-kuliner dengan klaim manfaat kesehatan harus diikuti oleh edukasi dan transparansi. Temuan ini sejalan dengan studi oleh Widnyani dan Aristayudha (2022), yang menyatakan bahwa persepsi negatif pada produk madu sering kali muncul dari visualisasi produk yang tidak sesuai serta asumsi konsumen yang tidak teredukasi.

d. Batik

Produk batik mencatat 24% sentimen negatif dan 18% netral, menjadikannya kategori dengan proporsi kepuasan paling rendah. Komentar negatif dominan muncul

pada perbedaan warna, bahan yang tipis, serta ukuran tidak sesuai deskripsi. Ulasan seperti “warnanya lebih pudar dari foto”, “bahan terlalu tipis”, atau “ukuran tidak sesuai chart” menggambarkan bahwa ekspektasi visual dan deskriptif memiliki peran besar dalam persepsi terhadap produk fashion. Ini menegaskan bahwa kualitas foto produk dan deskripsi yang akurat sangat krusial. Sesuai dengan penelitian Hoiriyyah et al. (2023), produk fashion seperti batik sangat rentan terhadap mismatch karena konsumen tidak dapat mencoba langsung, dan ketidaksesuaian kecil dapat berdampak besar pada keputusan pembelian berikutnya.

3.3 Pembahasan Pola Umum Sentimen

Sentimen Positif: Umumnya muncul pada produk yang memenuhi ekspektasi sensorik seperti rasa, aroma, dan visual. Kategori kuliner unggul dalam hal ini. Sentimen Netral: Biasanya berupa komentar pendek tanpa ekspresi emosi kuat, seperti “oke”, “sesuai”, “lumayan”. Netral sering muncul saat produk tidak memberikan kesan yang menonjol. Sentimen Negatif: Dipicu oleh ketidaksesuaian antara ekspektasi (dari deskripsi atau foto) dengan kenyataan. Produk non-kuliner memiliki potensi konflik lebih besar di sini.

3.4 Implikasi Hasil Penelitian

Kesesuaian Deskripsi Visual sangat penting, terutama pada produk non-kuliner seperti batik dan madu. Ketidaksesuaian warna, ukuran, atau tekstur menjadi pemicu utama sentimen negatif. Produk kuliner unggul karena pengalaman sensorik langsung. Rasa, aroma, dan packaging memainkan peran besar dalam persepsi positif. Ulasan dapat menjadi sumber insight strategis. Dengan pendekatan sederhana seperti klasifikasi leksikal, UMKM dapat memetakan persepsi pasar secara cepat tanpa harus bergantung pada survei.

Marketplace bukan hanya saluran jual, tetapi juga saluran riset pasar. Data ulasan yang tersedia secara publik dapat digunakan secara etis untuk perbaikan produk secara berkelanjutan (Rahmi & Dari, 2024; Nuraini et al., 2025).

CONCLUSION

4.1 Kesimpulan

Penelitian ini berhasil menerapkan pendekatan Natural Language Processing (NLP) berbasis leksikon untuk menganalisis sentimen konsumen terhadap empat kategori produk UMKM di Indonesia, yaitu keripik, sambal, madu hutan, dan batik, dengan memanfaatkan data ulasan dari Tokopedia dan Shopee. Hasil klasifikasi menunjukkan bahwa, produk kuliner keripik dan sambal secara umum mendapatkan sentimen positif dominan yaitu lebih dari 75%, yang berkaitan dengan rasa, kemasan, dan layanan penjual. Produk non-kuliner batik dan madu hutan, menunjukkan proporsi sentimen negatif lebih tinggi, dipicu oleh masalah kualitas, perbedaan warna/ukuran, serta ketidaksesuaian antara deskripsi dan kenyataan produk. Metode lexicon-based terbukti efektif dan efisien untuk konteks UMKM dengan keterbatasan teknologi, tanpa perlu model pembelajaran mesin kompleks. Validasi manual juga menunjukkan akurasi klasifikasi yang cukup tinggi (>90%), mendukung penggunaan metode ini dalam praktik bisnis sederhana. Secara umum, penelitian ini menegaskan pentingnya transparansi informasi produk, akurasi deskripsi, dan penggunaan data ulasan sebagai sumber insight strategis bagi pelaku UMKM di ranah digital.

4.2 Rekomendasi

Berdasarkan hasil temuan dan analisis, peneliti memberikan beberapa rekomendasi praktis dan akademis sebagai berikut, Untuk UMKM perbaiki deskripsi produk dan visual foto agar sesuai dengan kondisi aktual barang, terutama pada produk fashion dan herbal, Pantau secara rutin ulasan konsumen di marketplace untuk mengidentifikasi tren persepsi negatif dan melakukan tindakan korektif. Untuk pengembang sistem, kembangkan dashboard analisis sentimen berbasis lexicon sederhana yang dapat digunakan oleh UMKM tanpa pengetahuan teknis tinggi. Untuk penelitian selanjutnya, gunakan kombinasi pendekatan leksikon dan metode pembelajaran mesin seperti Naïve Bayes atau SVM, untuk meningkatkan akurasi klasifikasi. Tambahkan aspek analisis aspek (aspect-based sentiment analysis) agar lebih detail dalam memahami elemen spesifik seperti rasa, harga, atau pengemasan.

REFERENCES

- Agustiono, W., & Darmawan, A. K. (2023). E-Marketplace selection for Indonesian MSMEs: A sentiment analysis approach combining lexicon-based and support vector machine. 2023 IEEE 9th International Conference on Information Technology Systems and Innovation (ICITSI). <https://ieeexplore.ieee.org/document/10420382>
- Bhagaskara, B. A., Ismail, A., & Diena, H. E. (2023). Analisis Sentimen Produk Pada Bukalapak Menggunakan Lexicon-Based dan Multinomial Naïve Bayes. Jurnal Elektronik Ilmu Komputer. <https://www.researchgate.net/publication/392222317>
- Hoiriyah, H., Mardiana, H., & Walid, M. (2023). Lexicon-based and Naive Bayes Sentiment Analysis for Recommending the Best Marketplace Selection as a Marketing Strategy for MSMEs. Jurnal Pilar Nusa Mandiri, 19(2), 129–136. <https://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/4176>
- Kementerian Koperasi dan UKM. (2023). Laporan Tahunan KemenkopUKM 2023. <https://kemenkopukm.go.id>. <https://kemenkopukm.go.id/uploads/laporan>
- Muttakin, F., Andrika, N., & Salsabila, S. (2025). Sentiment Analysis of Shoe Product Reviews on Indonesian E-Commerce Platform Using Lexicon Based and Support Vector Machine. Jurnal Teknik Informatika (JUTIF), 7(1), 15–22. <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/3800>
- Nuraini, R. S., Krisbiantoro, D., & Imron, M. (2025). Analisis Sentimen Pengguna Aplikasi Agoda di Google Play Store Menggunakan Algoritma Naive Bayes. Journal of Information System Management (JOISM), 3(1), 45–52. <https://jurnal.amikom.ac.id/index.php/joism/article/view/2066>
- Purbaya, M. E., & Rakhmadani, D. P. (2022). Comparison of Kernel Support Vector Machines in Conducting Sentiment Analysis Review of Buying Chips on the Shopee E-Marketplace in Indonesian. 2022 International Conference on Informatics (ICOT). <https://ieeexplore.ieee.org/document/10017546>
- Rahmi, N. A., & Dari, R. W. (2024). Implementation of Natural Language Processing in Consumer Sentiment Analysis of Product Comments on the Marketplace. Jurnal Teknik Informatika (JUTIF), 6(2), 55–63. <http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/1666>
- Setyabudi, S., & Aryanny, E. (2025). Sentiment Analysis of Lazada Marketplace User Ratings with Naïve Bayes and Support Vector Machine Methods. Jurnal INOVTEK Polbeng, 11(1), 25–31. <https://jurnal.polbeng.ac.id/index.php/ISI/article/view/395>
- Sari, D. R., Matsaany, B., & Hamka, M. (2023). Aspect Extraction of E-Commerce and Marketplace Applications Using Word2Vec and WordNet Path. Jurnal Teknik Informatika (JUTIF), 5(4), 203–211. <http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/726>

- Vincent, V., & Tamba, I. F. U. (2022). Level of Employee Loyalty from Generation Z. Almana: Jurnal Manajemen dan Bisnis, 6(2), 249-257.
- Wahyudi, N. (2024). Sistem Analisis Sentimen Terhadap Produk Sunscreen pada Marketplace Shopee Menggunakan Support Vector Machine (SVM) [Skripsi, Universitas Islam Sultan Agung]. <http://repository.unissula.ac.id/id/eprint/37451>
- Widnyani, N. M., & Aristayudha, A. (2022). Sentiment Analysis Design of Product Reviews in the Marketplace Using Naïve Bayes Classifier Method; A Case Study in Tokopedia. Jurnal Ekonomi Digital, 4(3), 11–20. <https://www.ejournal.seaninstitute.or.id/index.php/Ekonomi/article/view/259>