TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

Analisis Sentimen Terhadap Ulasan Google Review Honda Malang Sekawan Motor Menggunakan IndoRoBERTa

Fera Dhea Sagita, M. Syauqi*, Risqy Siwi Pradini

Fakultas Sains dan Teknologi, Prodi Informatika, Institut Teknologi Sains dan Kesehatan RS DR. Soepraoen Kesdam V/BRW, Kota Malang, Indonesia
Email: [1]verasagita26@gmail.com, [2,*]haris@itsk-soepraoen.ac.id, [3]risqypradini@itsk-soepraoen.ac.id
Email Penulis Korespondensi: haris@itsk-soepraoen.ac.id

Abstrak–Ulasan pelanggan pada Google Review mengandung informasi penting mengenai kualitas layanan, namun jumlah data yang besar dan tidak terstruktur menyulitkan proses analisis secara manual. Penelitian ini bertujuan menganalisis sentimen ulasan pelanggan pada Google Review Honda Malang Sekawan Motor Singosari menggunakan model IndoRoBERTa melalui proses fine-tuning. Dataset diperoleh melalui proses web scraping, kemudian dilakukan preprocessing yang meliputi cleaning, case folding, normalisasi, tokenisasi, stopword removal, dan stemming. Pelabelan sentimen dilakukan menggunakan metode lexicon-based InSet sebagai pelabelan awal (weak labeling), kemudian dataset dibagi menjadi data latih dan data uji dengan rasio 80:20. Model digunakan untuk mengklasifikasikan sentimen ke dalam tiga kelas, yaitu positif, netral, dan negatif. Hasil pengujian menunjukkan bahwa model memperoleh nilai accuracy sebesar 0,94, precision sebesar 0,92, recall sebesar 0,94, dan F1-score sebesar 0,92. Model menunjukkan performa yang baik dalam mengklasifikasikan sentimen positif, namun masih mengalami keterbatasan dalam mengenali sentimen negatif akibat ketidakseimbangan distribusi data. Kontribusi penelitian ini adalah memberikan evaluasi penerapan fine-tuning IndoRoBERTa pada domain ulasan Google Review dealer otomotif berbahasa Indonesia yang memiliki karakteristik berbeda dari penelitian terdahulu yang banyak menggunakan data media sosial, sekaligus menganalisis pengaruh weak labeling berbasis InSet dan ketidakseimbangan kelas terhadap performa klasifikasi. Hasil penelitian dapat menjadi dasar pemanfaatan analisis sentimen untuk membantu pihak dealer memahami persepsi pelanggan dan mengidentifikasi aspek layanan yang perlu dipertahankan maupun diperbaiki. Evaluasi penelitian juga menunjukkan bahwa hasil klasifikasi masih merepresentasikan pola label yang dihasilkan InSet, sehingga penelitian selanjutnya perlu menggunakan anotasi manusia sebagai ground truth untuk meningkatkan validitas hasil.

Kata Kunci: Ulasan Pelanggan Analisis Sentimen; Fine-Tuning; Google Review; IndoRoBERTa; Weak Labeling

Abstract–Customer reviews on Google Review contain important information about service quality, but the large and unstructured amount of data makes manual analysis difficult. This research aims to analyze the sentiment of customer reviews on Google Review Honda Malang Sekawan Motor Singosari using the IndoRoBERTa model thru a fine-tuning process. The dataset was obtained thru web scraping, followed by preprocessing which included cleaning, case folding, normalization, tokenization, stopword removal, and stemming. Sentiment labeling is performed using the lexicon-based InSet method as the initial labeling (weak labeling), then the dataset is divided into training and testing data with an 80:20 ratio. The model is used to classify sentiment into three classes: positive, neutral, and negative. The test results show that the model achieved an accuracy of 0.94, precision of 0.92, recall of 0.94, and an F1-score of 0.92. The model shows good performance in classifying positive sentiment, but still has limitations in recognizing negative sentiment due to data distribution imbalance. The contribution of this research is to provide an evaluation of the application of fine-tuning IndoRoBERTa in the domain of Google Review reviews of Indonesian-language automotive dealers, which have different characteristics from previous studies that mostly used social media data, while also analyzing the influence of InSet-based weak labeling and class imbalance on classification performance. The research results can serve as a basis for utilizing sentiment analysis to help dealers understand customer perceptions and identify aspects of service that need to be maintained or improved. The evaluation of the research also shows that the classification results still represent the label patterns generated by InSet, so further research needs to use human annotation as ground truth to improve the validity of the results.

Keywords: Sentiment Analysis; Fine-Tuning; Google Review; IndoRoBERTa; Weak Labeling

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong perubahan signifikan dalam perilaku konsumen, serta memungkinkan konsumen untuk secara aktif menyampaikan dan berbagi pengalaman terhadap suatu layanan melalui media digital (Mukherjee, 2024). Salah satu platform ulasan online yang banyak digunakan adalah *Google Reviews*, yang memungkinkan pelanggan memberikan penilaian dan opini secara terbuka serta mudah diakses oleh publik (Fernandes, 2022). Dalam sektor otomotif, ulasan pelanggan menjadi faktor penting yang memengaruhi keputusan pembelian serta membentuk citra perusahaan (Chockalingam, 2023). Dealer dengan rating tinggi dan ulasan positif cenderung lebih dipercaya oleh calon konsumen dibandingkan dengan dealer dengan reputasi yang kurang baik, karena ulasan online berperan dalam membentuk kepercayaan dan persepsi konsumen terhadap suatu bisnis (Ahn & Lee, 2024).

Salah satu contoh nyata adalah Honda Malang Sekawan Motor Singosari, yang memiliki rating tinggi (4,8) pada *Google Reviews.* Jumlah ulasan yang besar ini menunjukkan tingginya interaksi pelanggan serta mencerminkan pentingnya peran ulasan daring dalam membentuk persepsi dan memengaruhi keputusan konsumen terhadap suatu layanan (Guo & Luo, 2023). Ulasan tersebut mengandung berbagai opini terkait kualitas pelayanan, fasilitas, harga, hingga pengalaman transaksi. Namun, jika tidak dianalisis secara sistematis, informasi penting yang terkandung dalam ulasan tersebut dapat menyebabkan pihak manajemen kehilangan insight penting untuk peningkatan layanan.

Permasalahan utama dalam pemanfaatan ulasan pelanggan terletak pada bentuk data yang tidak terstruktur serta mengandung informasi semantik yang kompleks, sehingga memerlukan teknik pengolahan teks untuk mengekstraknya (Liang, 2022). Banyak ulasan menggunakan bahasa Indonesia informal, singkatan, bahkan campuran bahasa daerah, yang

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

menyulitkan proses analisis secara manual. Selain itu, volume ulasan pelanggan yang besar serta sifatnya sebagai data tidak terstruktur menjadikan analisis secara manual kurang efisien (Lee, n.d.). Melimpahnya ulasan pelanggan di platform digital menunjukkan pentingnya kemampuan dalam mengelola dan memanfaatkan informasi tersebut secara efektif untuk mendukung pengambilan keputusan (Walther et al., 2026). Dalam praktiknya, pihak dealer sering kali hanya berfokus pada nilai rating secara umum tanpa memahami detail sentimen yang mendasarinya. Padahal, ulasan dengan rating yang sama dapat memiliki karakteristik ulasan yang sangat berbeda, misalnya dominasi keluhan pada aspek pelayanan atau kepuasan pada aspek tertentu.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan berbasis *Natural Language Processing (NLP)* untuk mengekstrak informasi yang relevan dan mendukung pengambilan keputusan (Berry, 2024). Analisis sentimen secara khusus dimanfaatkan untuk mengklasifikasikan opini pelanggan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Dalam beberapa tahun terakhir, metode berbasis deep learning semakin banyak diterapkan karena mampu memahami konteks kalimat serta keterkaitan antarkata dengan lebih baik. Jika dibandingkan dengan metode tradisional seperti *Naïve Bayes* dan *Support Vector Machine* (SVM), pendekatan ini dinilai memiliki kemampuan yang lebih efektif dalam menangani kompleksitas bahasa secara mendalam (Hidayat et al., 2025).

Salah satu model yang umum diterapkan dalam NLP adalah RoBERTa *(Robustly Optimized BERT Pretraining Approach)*, yang merupakan pengembangan dari BERT *(Bidirectional Encoder Representations from Transformers)* dengan peningkatan pada proses pelatihan dan representasi bahasa. Model RoBERTa memiliki berbagai keunggulan, di antaranya kemampuan memahami konteks bahasa secara dua arah, toleransi terhadap noise linguistik, serta kemampuan mentransfer pengetahuan dari proses *pretraining* ke korpus data yang lebih besar (Purwati, 2023). Selanjutnya, model tersebut dikembangkan menjadi IndoRoBERTa yang secara khusus dilatih menggunakan korpus berbahasa Indonesia sehingga memiliki kemampuan lebih baik dalam memahami konteks lokal, termasuk ragam bahasa informal yang sering muncul pada ulasan di *Google Review.* Melalui mekanisme *self-attention*, model ini dapat mengenali keterkaitan antarkata dalam sebuah kalimat secara lebih optimal.

Sejumlah penelitian terdahulu telah mengkaji analisis sentimen dengan menerapkan beragam metode. Salah satu penelitian yang memanfaatkan model berbasis *Recurrent Neural Network* (RNN), yaitu *Long Short-Term Memory* (LSTM), memperoleh hasil yang cukup baik dalam mengklasifikasikan sentimen terkait isu vaksinasi COVID-19 di Twitter dengan nilai akurasi sekitar 80,34% (Chintya Ayu Maharani, Budi Warsito, 2024). Sejumlah penelitian terdahulu telah mengeksplorasi penggunaan model BERT dan RoBERTa dalam analisis sentimen berbahasa Indonesia. Terdapat penelitian (Girsang, 2022) menggunakan model IndoBERT untuk mengklasifikasikan sentimen ke dalam dua kategori, yaitu positif dan negatif, dengan hasil *F1-score* sebesar 84%. Sementara itu, penelitian (Sandra et al., 2022) menerapkan model IndoRoBERTa untuk klasifikasi sentimen menghasilkan akurasi 91.73% dan *F1-Score* 86.42%. Namun, penelitian tersebut berfokus pada data Twitter yang memiliki karakteristik bahasa dan konteks berbeda dengan ulasan pelanggan pada Google Review. Sementara itu, penelitian (Al-zaelani et al., 2023) menganalisis ulasan mengenai produk motor Honda PCX dan Yamaha N-MAX pada Twitter menggunakan Naive Bayes. Penelitian tersebut telah berada pada domain otomotif, tetapi masih menggunakan algoritma machine learning konvensional dan sumber data Twitter, bukan ulasan pelanggan langsung pada *Google Review.*

Berdasarkan keempat penelitian tersebut, terdapat beberapa kesenjangan penelitian *(research gap)*. Pertama, penelitian terdahulu masih banyak menggunakan data Twitter atau media sosial sebagai objek analisis, sedangkan analisis sentimen pada *Google Review* dealer otomotif berbahasa Indonesia masih relatif terbatas. Kedua, penelitian yang menggunakan IndoRoBERTa sebelumnya lebih banyak diterapkan pada data media sosial, sehingga penerapannya pada ulasan pelanggan dealer otomotif dengan karakteristik bahasa informal, variasi panjang ulasan, dan konteks pelayanan masih perlu dievaluasi. Ketiga, penelitian terdahulu belum secara khusus mengkaji kombinasi *fine-tuning* IndoRoBERTa dengan pelabelan awal menggunakan InSet pada data *Google Review* dealer otomotif dengan tiga kelas sentimen. Keempat, penelitian sebelumnya belum secara spesifik membahas dampak ketidakseimbangan distribusi sentimen terhadap kemampuan model dalam mengenali kelas negatif pada domain tersebut. Oleh karena itu, penelitian ini mengisi kesenjangan tersebut dengan menerapkan *fine-tuning* IndoRoBERTa pada ulasan *Google Review* Honda Malang Sekawan Motor Singosari untuk klasifikasi tiga kelas sentimen, yaitu positif, netral, dan negatif. Pelabelan awal dilakukan menggunakan *InSet* sebagai pendekatan *weak labeling,* kemudian performa model dievaluasi menggunakan *accuracy, precision, recall, F1-score, dan confusion matrix.* Kontribusi utama penelitian ini bukan berupa pengembangan algoritma baru, melainkan evaluasi empiris penerapan IndoRoBERTa pada domain ulasan dealer otomotif berbahasa Indonesia serta analisis terhadap pengaruh weak labeling berbasis InSet dan ketidakseimbangan kelas terhadap hasil klasifikasi. Selain memberikan gambaran performa model pada domain yang masih terbatas diteliti, penelitian ini juga mengidentifikasi kelemahan model dalam mengenali kelas negatif sebagai dasar bagi penelitian selanjutnya untuk menerapkan anotasi manusia, penyeimbangan data, dan strategi peningkatan data.

Dengan demikian, penelitian ini bertujuan untuk menganalisis sentimen ulasan pelanggan menggunakan IndoRoBERTa serta mengevaluasi performanya pada domain ulasan layanan otomotif. Kontribusi penelitian ini tidak terletak pada pengembangan algoritma baru, melainkan pada evaluasi penerapan IndoRoBERTa pada dataset Google Review dealer otomotif berbahasa Indonesia yang memiliki karakteristik domain, distribusi sentimen, dan gaya bahasa yang berbeda dibandingkan dengan dataset media sosial yang umum digunakan. Selain itu, penelitian ini memberikan analisis terhadap dampak penggunaan pelabelan awal (*weak labeling*) berbasis InSet dan ketidakseimbangan distribusi kelas terhadap hasil klasifikasi sebagai dasar rekomendasi pengembangan penelitian selanjutnya. Dengan demikian, hasil

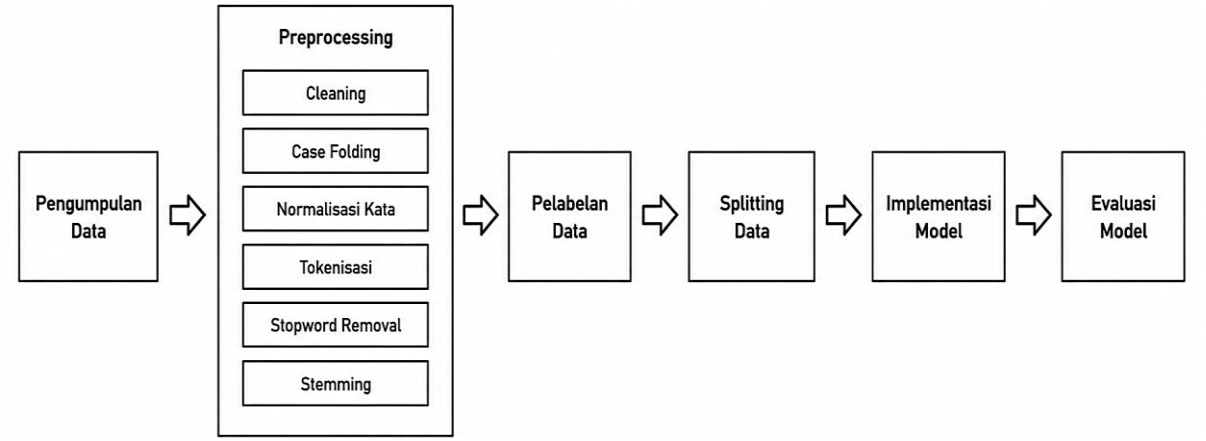
TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

penelitian diharapkan dapat memberikan masukan bagi pengelola dealer dalam memahami persepsi pelanggan sekaligus memperkaya kajian analisis sentimen pada domain ulasan layanan otomotif.

2. METODOLOGI PENELITIAN

Penelitian ini menerapkan pendekatan *Natural Language Processing (NLP)* untuk menganalisis sentimen ulasan pelanggan pada Google Review dealer Honda Malang Sekawan Motor Singosari. Objek penelitian berupa ulasan pelanggan yang dipublikasikan pada Google Review dan dikumpulkan melalui proses web scraping. Dataset yang diperoleh terdiri atas teks ulasan beserta informasi pendukung, kemudian melalui proses seleksi untuk menghilangkan data duplikat, data kosong, dan data yang tidak relevan sehingga diperoleh dataset yang layak digunakan dalam penelitian.

Tahapan penelitian dirancang secara sistematis mulai dari pengumpulan data, *preprocessing*, pelabelan sentimen, pembagian dataset, implementasi model IndoRoBERTa melalui proses *fine-tuning*, hingga evaluasi performa model. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1, yang menggambarkan hubungan antar setiap tahapan penelitian mulai dari proses akuisisi data hingga evaluasi hasil klasifikasi.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Studi ini memanfaatkan data berupa ulasan pelanggan yang diperoleh melalui *Google Reviews* pada dealer Honda Malang Sekawan Motor Singosari. Pengumpulan data dilakukan melalui metode web scraping dengan memanfaatkan Apify API sebagai alat bantu ekstraksi data. Data yang dikumpulkan meliputi teks ulasan, rating, serta informasi pendukung lainnya. Pengambilan data dilakukan terhadap seluruh ulasan pelanggan yang tersedia pada halaman *Google Reviews* dealer pada saat proses scraping dilakukan, sehingga data mencakup ulasan dari periode awal hingga ulasan terbaru pada bulan Mei 2026. Selanjutnya, dataset melalui tahap seleksi untuk memastikan kualitas data, seperti menghapus data ganda, data kosong, serta data yang tidak relevan, sehingga data yang digunakan lebih representatif untuk proses analisis sentimen.

2.2 Preprocessing

Tahapan preprocessing dilakukan dengan tujuan membersihkan serta mempersiapkan data sehingga siap untuk diolah oleh model. Data yang diperoleh dari proses scraping tidak dapat langsung digunakan untuk klasifikasi karena masih mengandung berbagai *noise.* Oleh karena itu, diperlukan tahap preprocessing untuk membersihkan dan menstrukturkan data sehingga siap digunakan dalam proses klasifikasi (Purnama, 2023). Proses ini meliputi:

1. Cleaning
   Tahapan ini bertujuan untuk menghapus berbagai karakter yang tidak relevan, seperti tanda baca, angka, simbol, dan emoji, sehingga data menjadi lebih rapi serta lebih mudah diolah pada proses selanjutnya (Haris et al., 2024).
2. Case Folding
   Pada tahap ini, semua huruf dalam teks ulasan diubah menjadi huruf kecil *(lowercase)* untuk menyeragamkan bentuk penulisan kata. Langkah tersebut bertujuan agar kata-kata yang memiliki arti sama tetapi berbeda penggunaan huruf kapital tetap dikenali sebagai kata yang identik (Egha Arya Affandi, Sumarno Sumarno, Uce Indahyanti, 2026).
3. Normalisasi Kata
   Sebagai bagian dari tahapan preprocessing dalam *Natural Language Processing* (NLP), proses normalisasi dilakukan dengan mengonversi kata tidak baku, singkatan, dan istilah slang ke dalam bentuk kata yang lebih standar. Menurut (Abidin & Pamungkas, 2025), normalisasi teks yang tidak baku, seperti singkatan dan ejaan yang tidak konsisten, penting untuk meningkatkan kinerja NLP, khususnya pada platform digital.
4. Tokenizing
   Dalam tahap ini, teks diuraikan menjadi unit-unit dasar berupa token, yaitu kata-kata individual. Proses tersebut memungkinkan sistem untuk mengidentifikasi dan memperlakukan setiap kata sebagai entitas yang berdiri sendiri dalam analisis (Egha Arya Affandi, Sumarno Sumarno, Uce Indahyanti, 2026).

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

5. Stopword Removal
   Pada proses ini dilakukan eliminasi terhadap kata-kata yang tergolong sebagai stopword, yaitu kata dengan frekuensi kemunculan tinggi namun memiliki nilai informasi yang rendah sehingga tidak berpengaruh signifikan terhadap hasil analisis (Fitri et al., 2020).
6. Stemming
   Proses yang dilakukan untuk mengubah kata dengan berbagai bentuk morfologis menjadi kata dasar. Tahapan ini dilakukan dengan menghapus imbuhan, baik berupa awalan, sisipan, akhiran, maupun gabungan dari beberapa imbuhan tersebut (Al-zaelani et al., 2023).

2.3 Pelabelan Data

Setelah melalui tahap pemrosesan, data ulasan selanjutnya diklasifikasikan ke dalam tiga label sentimen, yaitu positif, negatif, dan netral. Pemberian label dilakukan secara otomatis dengan memanfaatkan pendekatan berbasis leksikon menggunakan InSet *(Indonesian Sentiment Lexicon).* Metode ini bekerja dengan mencocokkan kata dalam ulasan dengan kamus sentimen berbahasa Indonesia yang memiliki skor polaritas, kemudian menjumlahkan skor tersebut untuk menentukan label akhir. Penggunaan InSet dipilih karena memiliki interpretabilitas tinggi, sehingga proses pelabelan dapat dijelaskan secara jelas berdasarkan skor kata. Selain itu, karena proses pelabelan dilakukan secara otomatis, tidak ada tahap verifikasi manual seperti pengukuran *inter-annotator agreement.* Pelabelan menggunakan InSet juga dipilih sebagai alternatif *low-cost annotation* yang konsisten, sementara proses *fine-tuning* IndoRoBERTa digunakan untuk menangkap konteks kalimat yang tidak terakomodasi dalam pendekatan leksikon. Kombinasi ini menjadi relevan terutama pada domain dengan keterbatasan data berlabel manual. Meskipun demikian, penggunaan InSet tetap membantu memastikan konsistensi hasil pelabelan serta meminimalkan kemungkinan bias subjektif (Abidin & Pamungkas, 2025).

Dalam proses pelabelan, diterapkan beberapa aturan *(rules lexicon)* untuk menentukan label sentimen secara sistematis. Setiap istilah yang terdapat pada ulasan terlebih dahulu dicocokkan dengan kamus InSet untuk memperoleh skor polaritas, baik positif maupun negatif. Selanjutnya, seluruh skor kata dalam satu ulasan dijumlahkan untuk mendapatkan skor sentimen total. Apabila skor total bernilai lebih dari nol (> 0), maka ulasan diklasifikasikan sebagai sentimen positif, sedangkan jika skor total kurang dari nol (< 0), maka dikategorikan sebagai sentimen negatif. Sementara itu, jika skor total sama dengan nol ($= 0$), ulasan diklasifikasikan sebagai sentimen netral. Kata-kata yang tidak terdapat dalam kamus leksikon dianggap memiliki skor nol sehingga tidak memengaruhi hasil akhir. Selain itu, dalam beberapa kasus, kata negasi seperti “tidak”, “bukan”, atau “kurang” diperhatikan karena dapat membalikkan polaritas kata yang mengikutinya, serta kata penguat seperti “sangat” atau “banget” dapat meningkatkan intensitas nilai sentimen (Koto, 2017).

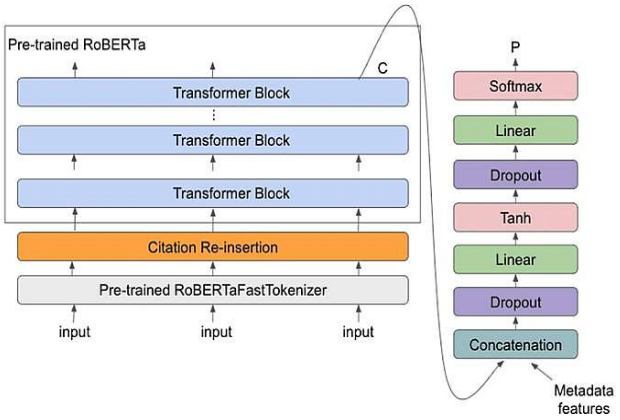
2.4 Splitting Dataset

Setelah proses pelabelan selesai dilakukan, dataset dibagi ke dalam dua kelompok, yaitu *training set* dan *testing set* dengan proporsi 80:20. *Training set* digunakan sebagai dasar pembelajaran model, sementara *testing set* dimanfaatkan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data baru yang belum pernah dikenali sebelumnya. Pembagian dataset dilakukan secara proporsional agar setiap kelompok data memiliki distribusi dan karakteristik yang tetap seimbang. Dengan demikian, model dapat memperoleh data yang cukup untuk proses pembelajaran sekaligus menyediakan data yang memadai untuk menguji kemampuan generalisasi model. Metode pembagian dengan proporsi 80:20 juga umum digunakan karena mampu meningkatkan efisiensi komputasi dan menghasilkan proses evaluasi model yang lebih stabil (Anshory et al., 2025).

2.5 Implementasi Model

Penelitian ini menerapkan arsitektur IndoRoBERTa, yaitu pengembangan model berbasis RoBERTa yang dirancang secara khusus untuk memahami karakteristik bahasa Indonesia. Model tersebut dibangun menggunakan arsitektur Transformer yang dikenal memiliki performa unggul dalam berbagai tugas NLP, termasuk pada proses klasifikasi teks (Dzakwan et al., 2025). Model ini menggunakan pendekatan *dynamic masking* selama proses pelatihan, serta memiliki arsitektur yang setara dengan BERT-Base, yaitu terdiri dari 12 lapisan *encoder,* dimensi *hidden* sebesar 768, dan 12 *attention heads,* dengan jumlah *parameter* sekitar 124 juta. Perbedaan utamanya terletak pada strategi pelatihan dan teknik masking yang diterapkan, di mana pola *masking* dapat berubah pada setiap batch data. Hal ini memungkinkan model untuk lebih adaptif dalam menangkap variasi konteks bahasa (Abidin & Pamungkas, 2025). Selain itu, IndoRoBERTa dilatih menggunakan korpus besar berbahasa Indonesia, sehingga mampu memahami struktur kalimat, makna kata, serta konteks linguistik yang lebih spesifik dibandingkan dengan model umum berbahasa Inggris. Kemampuan ini menjadikan IndoRoBERTa lebih efektif dalam menangani variasi bahasa sehari-hari, termasuk penggunaan bahasa informal yang sering ditemukan pada ulasan pengguna di platform seperti *Google Maps.* Dengan memanfaatkan model pretrained ini, proses pembelajaran menjadi lebih efisien karena model telah memiliki representasi bahasa yang kuat, sehingga pada penelitian ini hanya diperlukan proses fine-tuning untuk menyesuaikan model dengan tugas klasifikasi sentimen yang spesifik.

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558



Gambar 2. Siklus Pretrained Model RoBERTa (Hidayat et al., 2025)

Dari perspektif teknis, IndoRoBERTa menggunakan mekanisme self-attention untuk mengidentifikasi hubungan antar-token dalam urutan teks. Mekanisme ini dirumuskan dalam persamaan (1):

$$Attention\ (Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (1)$$

Dalam persamaan (1), nilai perhatian diperoleh dengan menghitung perkalian antara kueri (Q) dan kunci K^T, kemudian membaginya dengan 〖√d〗_k untuk menjaga skor tetap terkontrol. Hasil ini kemudian diberikan ke fungsi softmax untuk menghasilkan bobot perhatian, yang kemudian digunakan untuk menggabungkan nilai (V) (Roberta, n.d.). Hasil dari perhitungan attention pada persamaan (1) merepresentasikan tingkat kepentingan setiap token terhadap token lainnya dalam satu urutan teks. Dengan kata lain, setiap kata dalam kalimat tidak diproses secara terpisah, melainkan diperkaya dengan informasi dari kata-kata lain yang relevan. Bobot perhatian yang dihasilkan dari fungsi softmax menunjukkan seberapa besar kontribusi masing-masing token dalam membentuk representasi akhir suatu kata. Proses ini memungkinkan model untuk menangkap hubungan kontekstual yang kompleks, seperti dependensi jarak jauh antarkata, yang sulit dicapai oleh model berbasis sekuensial tradisional.

Selanjutnya, dalam arsitektur IndoRoBERTa, mekanisme ini tidak hanya dilakukan satu kali, tetapi juga melalui konsep *multi-head attention,* di mana beberapa proses *self-attention* dijalankan secara paralel. Setiap head akan mempelajari aspek hubungan yang berbeda antar-token, sehingga representasi yang dihasilkan menjadi lebih kaya dan informatif. Output dari setiap head kemudian digabungkan dan diproses lebih lanjut melalui *feed-forward network* pada setiap *transformer block.* Dengan pendekatan ini, IndoRoBERTa mampu memahami konteks bahasa Indonesia secara lebih mendalam, termasuk variasi makna kata yang bergantung pada konteks kalimat (Liu et al., 2019).

Tabel 1. Tabel Konfigurasi Model IndoRoBERTa

Parameter	Nilai
Metode	Fine-Tuning
Tokenizer	IndoRoBERTa
Max Sequence Length	128
Learning Rate	2e-5
Batch Size (Train)	16
Batch Size (Eval)	32
Jumlah Epoch	5
Optimizer	AdamW

Pada tahap implementasi model, penelitian ini menggunakan model pretrained berbasis arsitektur Transformer, yaitu IndoRoBERTa dengan model spesifik "w11wo/indonesian-roberta-base-sentiment-classifier". Pada penelitian ini, proses yang dilakukan bukan pelatihan model dari awal *(training from scratch),* melainkan *fine-tuning* model *pretrained* agar dapat menyesuaikan pola sentimen pada dataset ulasan *Google Reviews* yang digunakan. Konfigurasi parameter model ditampilkan pada Tabel 1. Sebelum proses fine-tuning, dataset terlebih dahulu melalui tahap preprocessing dan pelabelan sentimen menggunakan metode InSet sehingga menghasilkan tiga kelas sentimen, yaitu positif, netral, dan negatif. Selanjutnya, data dibagi menjadi data latih dan data uji dengan rasio 80:20. Teks ulasan kemudian diproses menggunakan tokenizer bawaan IndoRoBERTa dengan panjang maksimum 128 token. Proses *fine-tuning* dilakukan menggunakan *learning rate* sebesar 2e-5, *batch size* 16 untuk data latih dan 32 untuk data uji, serta jumlah *epoch* sebanyak 5. Setelah proses pelatihan selesai, model digunakan untuk memprediksi sentimen data uji berdasarkan probabilitas tertinggi dari masing-masing kelas sentimen.

2.5 Evaluasi Model

Setelah tahap pelatihan, model IndoRoBERTa yang telah disesuaikan selanjutnya diuji untuk mengevaluasi tingkat kinerjanya dalam mendeteksi kesalahan pada teks bahasa Indonesia. Proses evaluasi menggunakan metrik standar untuk

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

klasifikasi, yaitu *accuracy, presisi, recall*, dan *F1-score,* yang secara terpadu memberikan gambaran menyeluruh terkait kemampuan model dalam mengidentifikasi dan membedakan kategori kesalahan, terutama pada data ulasan *Google Review* (Opitz, 2024; Sukma et al., 2025).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - Score = 2\,x\,\frac{Precision\ x\ Recall}{Precision+Recall} \quad (5)$$

Melalui confusion matrix, hasil prediksi model dapat dianalisis berdasarkan empat komponen utama, yaitu *True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN).* Penggunaan kombinasi metrik ini memungkinkan evaluasi kinerja model dilakukan secara menyeluruh, khususnya dalam menangani variasi sentimen dalam teks bahasa Indonesia di media sosial. Keempat metrik tersebut saling melengkapi dalam mengevaluasi kinerja model. Accuracy mengukur ketepatan secara keseluruhan, namun kurang optimal pada data yang tidak seimbang. Oleh karena itu, *precision dan recall* digunakan untuk menilai ketepatan dan kemampuan model dalam mengenali setiap kelas, sedangkan *F1-score* memberikan keseimbangan antara keduanya. Kombinasi metrik ini memungkinkan evaluasi model dilakukan secara lebih komprehensif dan akurat.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dan pembahasan berdasarkan seluruh tahapan metodologi yang telah dilakukan, mulai dari pengumpulan data, analisis data eksploratif, preprocessing, pembagian dataset, implementasi model, hingga evaluasi dan analisis hasil. Temuan yang diperoleh dianalisis untuk menilai kinerja model IndoRoBERTa dalam mengklasifikasikan sentimen ulasan pelanggan pada platform *Google Reviews.*

3.1 Hasil Pengumpulan Data

Berdasarkan hasil proses pengumpulan data yang telah dilakukan, diperoleh sebanyak 1.198 data ulasan dari *Google Reviews* melalui proses scraping. Proses ini dilakukan secara otomatis untuk mengekstraksi data ulasan pelanggan yang tersedia secara publik, sehingga memungkinkan pengumpulan data dalam jumlah yang cukup besar dalam waktu yang relatif singkat. Selain itu, data yang dikumpulkan berasal dari berbagai pengguna dengan latar belakang yang berbeda, sehingga mengandung variasi bahasa, gaya penulisan, serta tingkat kepuasan pelanggan yang beragam. Hal ini menjadikan dataset lebih representatif dalam menggambarkan kondisi nyata persepsi pelanggan terhadap layanan yang diberikan. Namun demikian, data yang diperoleh masih bersifat mentah *(raw data)* dan belum terstruktur dengan baik, sehingga memerlukan tahap preprocessing sebelum digunakan dalam proses analisis. Seluruh data yang telah dikumpulkan kemudian disimpan dan dirapikan dalam format CSV untuk memudahkan proses pengolahan pada tahap selanjutnya. Adapun struktur kolom dalam dataset ditunjukkan pada Tabel 2.

Tabel 2. Struktur Dataset

Title	Stars	Name	Text
Honda Malang Sekawan Motor Singosari #Honda Malang	5	Kirania	pelayanan ramah dan cepat 😍 👍🏻 👍🏻 Terimakasih
Honda Malang Sekawan Motor Singosari #Honda Malang	5	Jeon Wonwoo	Baru pertama kali service pelayanannya oke
Honda Malang Sekawan Motor Singosari #Honda Malang	5	tongek gaming	Pelayanan bagus, penjelasan yg cukup jelas dan baik
Honda Malang Sekawan Motor Singosari #Honda Malang	5		Prosesnya cepat dan mudah 👍
Honda Malang Sekawan Motor Singosari #Honda Malang	4	Ranisa Duwi Iffa Zuhd	Good response, detailed full check service, 👍 👍

3.2 Hasil Preprocessing

Hasil dari tahap preprocessing menunjukkan bahwa ulasan yang sebelumnya tidak terstruktur berhasil diproses menjadi data yang lebih rapi, konsisten, dan siap dimanfaatkan untuk analisis sentimen. Proses cleaning dilakukan dengan menghapus karakter yang tidak relevan seperti tanda baca, angka, simbol, serta tautan (URL), sehingga mampu mengurangi *noise* pada data. Selanjutnya, case folding mengubah seluruh huruf menjadi huruf kecil untuk menyeragamkan penulisan kata dan menghindari duplikasi makna akibat perbedaan kapitalisasi. Tahap normalisasi kata berperan dalam mengonversi kata tidak baku, singkatan, serta bahasa slang ke dalam bentuk baku agar lebih mudah

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

dipahami oleh model. Proses tokenisasi kemudian memecah kalimat menjadi unit kata, diikuti dengan stopword removal untuk menghilangkan kata-kata umum yang tidak memiliki makna signifikan terhadap sentimen. Terakhir, stemming mengubah kata ke bentuk dasarnya sehingga variasi morfologis dapat disederhanakan dan dimensi fitur dapat disederhanakan.

Selain itu, proses preprocessing juga berkontribusi dalam mengurangi ambiguitas makna kata, meningkatkan konsistensi representasi teks, serta memperkecil kompleksitas data yang akan diproses oleh model. Tahapan ini membantu model untuk lebih fokus pada kata-kata yang memiliki bobot informasi penting terkait sentimen. Dengan demikian, kualitas data menjadi lebih optimal dan berpengaruh positif terhadap kinerja model dalam mengklasifikasikan sentimen secara lebih akurat dan stabil. Adapun hasil keseluruhan dapat dilihat pada Tabel 3 yang disajikan.

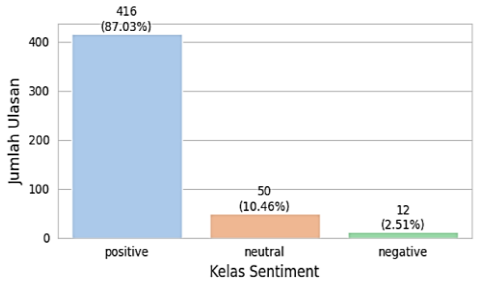
Tabel 3. Hasil Preprocessing Data

Tahapan	Hasil Preprocessing
Dataset Awal	Pelayanan cepat. Orangnya juga ramah dan sangat amanah. Terimakasih sekawan motor Singosari 👍
Cleaning	Pelayanan cepat Orangnya juga ramah dan sangat amanah Terimakasih sekawan motor Singosari
Case Folding	pelayanan cepat orangnya juga ramah dan sangat amanah terimakasih sekawan motor singosari
Normalisasi Kata	pelayanan cepat orangnya juga ramah dan sangat amanah terimakasih sekawan motor singosari
Tokenizing	['pelayanan', 'cepat', 'orangnya', 'juga', 'ramah', 'dan', 'sangat', 'amanah', 'terimakasih', 'sekawan', 'motor', 'singosari']
Stopword Removal	['pelayanan', 'cepat', 'orangnya', 'ramah', 'amanah', 'terimakasih', 'sekawan', 'motor', 'singosari']
Stemming	layan cepat orang ramah amanah terimakasih kawan motor singosari

3.3 Hasil Pelabelan Data

Setelah tahap preprocessing, jumlah data hasil scraping yang semula sebanyak 1.198 ulasan mengalami pengurangan akibat adanya nilai kosong *(missing values)* pada kolom teks serta proses pembersihan data. Data yang digunakan pada tahap akhir sebanyak 478 ulasan. Selanjutnya, data tersebut diberi label sentimen menggunakan metode berbasis leksikon, yaitu InSet *(Indonesian Sentiment Lexicon),* dengan cara mencocokkan setiap kata pada ulasan ke dalam kamus sentimen dan menjumlahkan skor polaritasnya untuk menentukan kategori sentimen. Dalam penelitian ini, penentuan label sentimen didasarkan pada total skor polaritas yang diperoleh dari setiap ulasan, dengan ketentuan sentimen positif diberikan apabila total skor > 0, sentimen negatif diberikan apabila total skor < 0, dan sentimen netral diberikan apabila total skor = 0. Sebagai contoh, ulasan “pelayanan the best” diklasifikasikan sebagai sentimen positif karena mengandung kata yang memiliki polaritas positif, seperti “best”, sehingga menghasilkan total skor positif. Pada ulasan “banyak macam motor”, tidak terdapat kata yang memiliki polaritas positif maupun negatif yang dominan dalam kamus leksikon, sehingga total skor bernilai 0 dan dikategorikan sebagai sentimen netral. Sementara itu, ulasan “pelayanan lama dan mengecewakan” diklasifikasikan sebagai sentimen negatif karena kata “lama” dan “mengecewakan” memiliki polaritas negatif yang menghasilkan total skor negatif. Pendekatan leksikon memungkinkan proses penentuan label lebih mudah dijelaskan karena didasarkan pada skor kata yang jelas. Selain itu, metode ini tidak memerlukan data latih tambahan, sehingga lebih efisien dan cocok digunakan sebagai baseline labeling, terutama ketika ketersediaan data berlabel masih terbatas. InSet juga dirancang khusus untuk bahasa Indonesia, sehingga mampu menangkap nuansa kata lokal dengan lebih relevan dibandingkan dengan leksikon umum.

Hasil pelabelan menunjukkan bahwa sebagian besar ulasan berada pada kategori positif. Dari total data, sebanyak 416 ulasan (87,03%) termasuk dalam sentimen positif, 50 ulasan (10,46%) tergolong netral, dan 12 ulasan (2,51%) termasuk dalam sentimen negatif. Dominasi sentimen positif ini mengindikasikan bahwa mayoritas pelanggan memberikan tanggapan yang baik terhadap layanan dealer Honda Malang Sekawan Motor Singosari. Meskipun demikian, keberadaan ulasan dengan sentimen negatif dan netral tetap menjadi perhatian penting karena mencerminkan adanya aspek layanan yang masih perlu diperbaiki. Informasi ini dapat dimanfaatkan oleh pihak manajemen sebagai bahan evaluasi untuk meningkatkan kualitas pelayanan secara keseluruhan.



Gambar 3. Distribusi Pelabelan Sentimen

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

Berdasarkan visualisasi *wordcloud* pada Gambar 4, terlihat perbedaan karakteristik kata yang muncul pada ulasan sentimen positif dan negatif. Pada *wordcloud* sentimen positif, kata-kata seperti "ramah", "pelayanan", "bagus", "nyaman", "terimakasih", dan "cepat" muncul dengan ukuran yang lebih dominan, yang menunjukkan bahwa pelanggan merasa puas terhadap kualitas layanan, keramahan staf, serta kenyamanan yang diberikan. Hal ini mengindikasikan bahwa aspek pelayanan menjadi faktor utama yang berkontribusi terhadap sentimen positif pelanggan. Sebaliknya, pada wordcloud sentimen negatif, kata-kata seperti "lama", "antri", "kurang", "tidak", dan "stok" lebih sering muncul. Kemunculan kata-kata tersebut menunjukkan adanya keluhan pelanggan terkait waktu pelayanan yang lambat, proses antrean, serta keterbatasan informasi atau ketersediaan produk. Dengan demikian, dapat disimpulkan bahwa perbedaan kata yang paling banyak muncul pada setiap kategori sentimen memberikan gambaran yang jelas mengenai aspek layanan yang perlu dipertahankan (pada sentimen positif) serta aspek yang perlu diperbaiki (pada sentimen negatif).

Gambar 4. Wordcloud Sentimen Positif dan Negatif

3.4 Hasil Splitting Dataset

Berdasarkan proses pembagian dataset dengan rasio 80:20, diperoleh data latih *(training set)* sebanyak 382 data dan data uji *(testing set)* sebanyak 96 data. Sebagian besar data dialokasikan sebagai data latih untuk memastikan model memiliki cukup informasi dalam proses pembelajaran, sedangkan sisanya digunakan sebagai data uji untuk mengevaluasi kinerja model. Dalam proses pelatihan, data uji juga digunakan sebagai data evaluasi sementara (validasi) pada setiap epoch untuk memantau performa model. Meskipun demikian, evaluasi akhir tetap dilakukan menggunakan data uji untuk mengukur kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

3.5 Hasil Implementasi Model

Pada tahap implementasi, penelitian ini berhasil menerapkan IndoRoBERTa melalui metode *fine-tuning* untuk mengklasifikasikan sentimen pada ulasan *Google Reviews.* Model yang digunakan berasal dari *pretrained* model "w11wo/indonesian-roberta-base-sentiment-classifier" yang sebelumnya telah melalui proses pelatihan dengan kumpulan data bahasa Indonesia dalam jumlah besar, sehingga mampu memahami konteks bahasa Indonesia secara lebih baik. Dalam penelitian ini, proses pelatihan tidak dilakukan dari awal *(training from scratch),* melainkan dengan menyesuaikan bobot model terhadap dataset ulasan yang digunakan. Sebelum proses pelatihan, dataset hasil preprocessing dan pelabelan sentimen dipisahkan ke dalam data latih dan data uji dengan perbandingan 80:20. Dari total 478 data, sebanyak 382 data digunakan sebagai data *training* dan 96 data sebagai data testing. Selanjutnya, teks ulasan diproses menggunakan tokenizer bawaan IndoRoBERTa dengan panjang maksimum 128 token agar dapat diubah menjadi representasi numerik yang dapat dipahami model. Proses *fine-tuning* dilakukan menggunakan *learning rate* sebesar 2e-5, *batch size* 16 untuk data latih, *batch size* 32 untuk data uji, serta jumlah epoch sebanyak 5. Selama proses *fine-tuning,* dilakukan evaluasi pada setiap *epoch* untuk memantau perkembangan performa model. Hasil *fine-tuning* model pada setiap *epoch* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Fine-Tuning Model Per Epoch

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.633800	0.331090	0.916667	0.903080	0.916667	0.894751
2	0.258298	0.315633	0.916667	0.903080	0.916667	0.894751
3	0.184599	0.366289	0.916667	0.903080	0.916667	0.894751
4	0.134905	0.313929	0.937500	0.920833	0.937500	0.922953
5	0.094398	0.342713	0.937500	0.920833	0.937500	0.922953

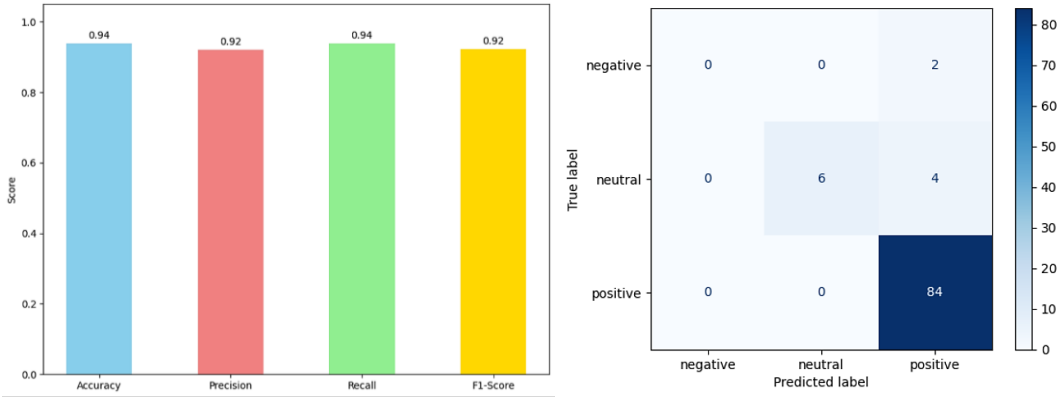
Berdasarkan hasil tersebut, nilai training loss menunjukkan tren penurunan secara bertahap dari *epoch* pertama hingga *epoch* kelima, yang menunjukkan bahwa model mampu mempelajari pola sentimen pada dataset dengan baik. Selain itu, performa model juga mengalami peningkatan, di mana nilai *accuracy* dan *F1-score* mencapai hasil terbaik pada *epoch* keempat dan kelima. Setelah proses *fine-tuning* selesai, model berhasil menghasilkan prediksi sentimen ke dalam tiga kelas, yaitu positif, netral, dan negatif berdasarkan nilai probabilitas tertinggi *(argmax)* dari masing-masing kelas sentimen.

3.6 Hasil Evaluasi Model

Tahap evaluasi model merupakan bagian penting dalam penelitian ini guna mengevaluasi seberapa baik model yang telah dibangun tersebut mampu melakukan klasifikasi sentimen secara akurat, konsisten, serta mampu melakukan generalisasi

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

terhadap data yang belum pernah dilihat sebelumnya. Evaluasi dilakukan dengan menggunakan beberapa metrik, yaitu *accuracy, precision, recall,* dan *F1-score,* yang masing-masing memberikan gambaran mengenai tingkat ketepatan prediksi, kemampuan model dalam mengenali setiap kelas, serta keseimbangan antara prediksi benar dan kesalahan. Selain itu, evaluasi juga didukung dengan visualisasi *confusion matrix* untuk melihat distribusi hasil prediksi terhadap label aktual pada setiap kelas sentimen, sehingga dapat diketahui kelebihan dan kekurangan model sebagai dasar untuk analisis lebih lanjut dan perbaikan pada tahap berikutnya.



(a) (b)

Gambar 5. (a) Performance Metrics IndoRoBERTa, (b) Confusion Matrix IndoRoBERTa

Berdasarkan Gambar 5(a), ditampilkan hasil evaluasi kinerja model dalam bentuk grafik batang. Nilai *accuracy* yang diperoleh sebesar 0,94, yang menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan sebesar 94%. Nilai *precision* sebesar 0,92 menunjukkan bahwa prediksi yang dihasilkan model memiliki tingkat ketepatan yang baik dalam mengidentifikasi kelas yang benar. Selain itu, nilai *recall* sebesar 0,94 mengindikasikan bahwa model mampu mengenali sebagian besar data aktual dengan baik. Nilai *F1-score* sebesar 0,92 menunjukkan keseimbangan yang cukup baik antara *precision dan recall,* sehingga model dapat dikatakan memiliki performa yang stabil dalam melakukan klasifikasi sentimen.

Selanjutnya, Gambar 5(b) menampilkan *confusion matrix* yang menunjukkan distribusi label aktual dan prediksi model. Model mampu mengklasifikasikan kelas positif dengan sangat baik, ditunjukkan oleh 84 prediksi benar tanpa kesalahan klasifikasi. Lalu, pada kelas netral, sebanyak 6 data berhasil diprediksi dengan benar dan 4 data lainnya salah diklasifikasikan sebagai positif. Hasil tersebut menunjukkan bahwa model masih cenderung mengklasifikasikan sebagian data netral ke dalam kelas positif. Pada kelas negatif, seluruh data belum dapat diprediksi dengan benar karena 2 data negatif diklasifikasikan sebagai positif. Kondisi ini menandakan bahwa model masih kurang optimal dalam mengenali sentimen negatif dan lebih dominan terhadap kelas positif. Secara keseluruhan, performa model sangat baik pada kelas positif, cukup baik pada kelas netral, namun masih rendah pada kelas negatif. Hal ini diduga dipengaruhi oleh ketidakseimbangan data, karena jumlah data negatif lebih sedikit sehingga pola sentimennya kurang dipelajari secara optimal oleh model. Walaupun model memiliki performa yang tinggi, analisis terhadap kesalahan klasifikasi tetap diperlukan untuk mengetahui keterbatasannya dalam memahami pola sentimen. Kesalahan prediksi dipengaruhi oleh beberapa faktor, seperti ketidakseimbangan data yang membuat model cenderung bias terhadap kelas positif. Selain itu, penggunaan bahasa tidak baku, singkatan, dan beragam ekspresi pada ulasan pelanggan juga menjadi tantangan bagi model dalam memahami konteks secara tepat.

Meskipun telah dilakukan tahap preprocessing, beberapa makna implisit dalam kalimat masih sulit ditangkap secara optimal oleh model. Sebagai contoh, pada ulasan “pelayanan lama tapi staf ramah”, label sebenarnya adalah netral, namun model memprediksi sebagai positif. Hal ini terjadi karena model lebih menangkap kata “ramah” yang memiliki polaritas positif dibandingkan dengan kata “lama” yang bersifat negatif. Contoh lainnya, pada ulasan “nunggu lama banget, tapi akhirnya dilayani dengan baik”, label sebenarnya adalah negatif, namun diprediksi sebagai positif. Kondisi ini menunjukkan bahwa model cenderung fokus pada kata dengan bobot sentimen yang lebih kuat, terutama pada bagian akhir kalimat seperti “baik”, sehingga kurang mempertimbangkan keseluruhan konteks. Selain itu, kesalahan juga dapat terjadi pada kalimat yang bersifat ambigu atau tidak eksplisit, seperti “ya begitulah pelayanannya”, yang dapat diinterpretasikan sebagai netral oleh manusia, namun sulit dipahami secara jelas oleh model. Oleh karena itu, meskipun model memiliki performa yang baik secara umum, masih diperlukan upaya peningkatan, seperti penyeimbangan data atau penambahan variasi data latih, agar model mampu mengklasifikasikan seluruh kategori sentimen secara lebih seimbang dan akurat.

3.7 Pembahasan

Hasil penelitian menunjukkan bahwa model IndoRoBERTa melalui proses *fine-tuning* memperoleh *accuracy* sebesar 0,94, *precision* sebesar 0,92, *recall* sebesar 0,94, dan *F1-score* sebesar 0,92. Hasil tersebut menunjukkan bahwa IndoRoBERTa mampu melakukan klasifikasi sentimen dengan baik pada ulasan *Google Review* Honda Malang Sekawan Motor Singosari. Jika dibandingkan dengan penelitian (Girsang, 2022) yang menggunakan IndoBERT dan memperoleh


TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

F1-score sebesar 84%, serta (Sandra et al., 2022) yang menggunakan IndoRoBERTa pada data Twitter dengan *accuracy* 91,73% dan F1-score 86,42%, hasil penelitian ini menunjukkan performa yang kompetitif. Perbedaan hasil tersebut dapat dipengaruhi oleh karakteristik dataset, domain ulasan, proses *fine-tuning*, serta distribusi kelas yang digunakan pada masing-masing penelitian. Pada penelitian (Al-zaelani et al., 2023) analisis sentimen produk Honda PCX dan Yamaha N-MAX menggunakan algoritma Naïve Bayes pada data Twitter memperoleh *accuracy* masing-masing sebesar 79% dan 72%. Sementara itu, penelitian (Chintya Ayu Maharani, Budi Warsito, 2024) menggunakan LSTM untuk menganalisis sentimen vaksinasi COVID-19 pada Twitter dan memperoleh *accuracy* sekitar 80,34%. Dibandingkan dengan kedua penelitian tersebut, hasil penelitian ini menunjukkan performa klasifikasi yang lebih tinggi. Namun, perbandingan tersebut tidak dapat diartikan sebagai keunggulan mutlak karena masing-masing penelitian menggunakan dataset, domain, jumlah data, dan metode pelabelan yang berbeda. Keunggulan hasil penelitian ini lebih menunjukkan bahwa IndoRoBERTa memiliki potensi yang baik untuk menangkap konteks bahasa pada ulasan *Google Review* berbahasa Indonesia.

Meskipun memperoleh nilai *accuracy* yang tinggi, hasil *confusion matrix* menunjukkan bahwa model masih mengalami kesulitan dalam mengenali sentimen negatif. Dari 478 data yang digunakan, sebanyak 416 data atau 87,03% merupakan sentimen positif, 50 data atau 10,46% netral, dan hanya 12 data atau 2,51% negatif. Kondisi tersebut menyebabkan model cenderung lebih baik dalam mengenali kelas mayoritas dibandingkan kelas negatif. Temuan ini menjadi salah satu pembeda penting dari penelitian terdahulu karena penelitian ini tidak hanya mengevaluasi performa keseluruhan model, tetapi juga menunjukkan keterbatasan model pada kelas minoritas. Selain itu, penggunaan InSet sebagai *weak labeling* menyebabkan hasil evaluasi masih bergantung pada kualitas label yang dihasilkan oleh leksikon, sehingga belum sepenuhnya merepresentasikan sentimen yang telah divalidasi oleh manusia. Dengan demikian, hasil penelitian menunjukkan bahwa penerapan IndoRoBERTa pada domain *Google Review* dealer otomotif memberikan performa yang baik, tetapi peningkatan kualitas anotasi dan penyeimbangan data masih diperlukan untuk memperoleh klasifikasi yang lebih seimbang pada seluruh kelas sentimen.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penerapan metode *lexicon-based InSet* sebagai pendekatan *weak labeling* yang dikombinasikan dengan model IndoRoBERTa melalui proses *fine-tuning* mampu melakukan analisis sentimen terhadap ulasan pelanggan pada *Google Review* Honda Malang Sekawan Motor Singosari dengan performa klasifikasi yang baik. Model mampu mengenali sentimen positif dan netral secara efektif, meskipun kemampuan mengenali sentimen negatif masih dipengaruhi oleh ketidakseimbangan distribusi data. Kontribusi penelitian ini terletak pada evaluasi penerapan IndoRoBERTa pada domain ulasan *Google Review* dealer otomotif berbahasa Indonesia serta analisis penggunaan InSet sebagai *weak labeling* dan pengaruh ketidakseimbangan kelas terhadap hasil klasifikasi. Hasil penelitian juga menunjukkan bahwa pendekatan berbasis Transformer berpotensi membantu manajemen dealer dalam memahami persepsi pelanggan dan menentukan aspek layanan yang perlu dipertahankan maupun diperbaiki. Namun, pelabelan menggunakan InSet tanpa anotasi manual menyebabkan hasil evaluasi masih merepresentasikan kemampuan model dalam mempelajari label leksikon, sedangkan ketidakseimbangan data menyebabkan sentimen negatif belum dikenali secara optimal. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan anotasi manual sebagai *ground truth*, teknik penyeimbangan data, serta mengeksplorasi model Transformer lainnya guna memperoleh hasil analisis yang lebih akurat dan objektif.

REFERENCES

Abidin, M. I., & Pamungkas, E. W. (2025). Analisis Sentimen Terhadap Timnas Indonesia di Piala Asia 2023 dengan Model Transformer Berbahasa Indonesia. *Jurnal Teknologi Dan Sistem Informasi Univrab*, *10*(2), 482–496. https://doi.org/https://doi.org/10.36341/rabit.v10i2.6142

Ahn, Y., & Lee, J. (2024). The Impact of Online Reviews on Consumers ' Purchase Intentions : Examining the Social Influence of Online Reviews , Group Similarity , and Self-Construal. *Journal of Theorectical and Applied Electronic Commerce Research*, *19*(2), 1060–1078. https://doi.org/https://doi.org/10.3390/jtaer19020055

Al-zaelani, R. S. A., Ramadhan, Y. R., & Komara, M. A. (2023). Analisis Sentimen Review Produk Motor Honda PCX dan Yamaha N-MAX Pada Twitter Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(3), 1812–1816. https://doi.org/https://doi.org/10.36040/jati.v7i3.7008

Anshory, M. N., Mazdadi, M. I., Saragih, T. H., & Saputro, S. W. (2025). Application of Adaboost Algorithm with SMOTE and Optuna Techniques in Sleep Disorder Classification. *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, *2*, 415–426. https://doi.org/https://doi.org/10.35882/ijeeemi.v7i2.99

Berry, S. (2024). *Consumer lying in online reviews : recent evidence*. *2010*, 1–16. https://doi.org/https://doi.org/10.48550/arXiv.2405.12743

Chintya Ayu Maharani, Budi Warsito, R. S. (2024). Analisis sentimen vaksin covid-19 pada twitter menggunakan recurrent neural network (rnn) dengan algoritma long short-term memory (lstm) 1,2,3. *Jurnal Gaussian*, *12*, 403–413. https://doi.org/10.14710/j.gauss.12.3.403-413

Chockalingam, D. (2023). Impact of Online Reviews and Ratings on Vehicle Purchase Decisions. *International Journal*

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1566-1576
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10558

for Multidisciplinary Research (IJFMR), *5*(4), 1–4.

Dzakwan, M., Lazuardi, B., & Fatyanosa, T. N. (2025). Klasifikasi Emosi Multikelas Berbasis Teks Bahasa Indonesia Menggunakan IndoRoBERTa. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *9*(9), 1–9.

Egha Arya Affandi, Sumarno Sumarno, Uce Indahyanti, Y. F. (2026). *Analisis Sentimen Ulasan Masyarakat Aplikasi Brompit Menggunakan Algoritma Naive Bayes dan SVM*. *10*(1), 1160–1167. https://doi.org/https://doi.org/10.36040/jati.v10i1.16955

Fernandes, S. (2022). Measuring the Impact of Online Reviews on Consumer Purchase Decisions – A Scale Development Study Measuring the Impact of Online Reviews on Consumer Purchase Decisions – A Scale Development Study. *Elsevier*, 0–34. https://www.sciencedirect.com/science/article/pii/S096969892200159X

Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). *Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes , Random Forest Dan Support Vector Machine*. *18*(1), 71–80.

Girsang, A. S. (2022). Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method. *International Journal of Engineering Trends and Technology*, *70*(12), 281–288. https://doi.org/https://doi.org/10.14445/22315381/IJETT-V70I12P226

Guo, W., & Luo, Q. (2023). Journal of Retailing and Consumer Services Investigating the impact of intelligent personal assistants on the purchase intentions of Generation Z consumers : The moderating role of brand credibility. *Journal of Retailing and Consumer Services*, *73*(April).

Haris, M., Suharso, A., Nurkifli, E. H., Informatika, P. S., Karawang, U. S., & Timur, T. (2024). Analisis Sentimen Pada Game Efootball di Google Play Store Menggunakan Algoritma IndoBERT. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(6), 12108–12121. https://doi.org/https://doi.org/10.36040/jati.v8i6.11810

Hidayat, J. J., Setyowati, C., Dikaisa, M., Amin, I., Bimasakti, K., & Werdana, A. P. (2025). Deep Learning-based Sentiment Analysis of Public Comments on Military Education Using RoBERTa Algorithm and Rule-Based Hybird Parameters. *Jurnal Media Computer Science*, *4*(2), 277–292. https://doi.org/https://doi.org/10.37676/jmcs.v4i2.8769

Koto, F. (2017). InSet Lexicon : Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs. *IEEE*, 391–394. https://doi.org/10.1109/IALP.2017.8300625

Lee, D. (n.d.). *Factors that affect consumer trust in product quality: a focus on online reviews and shopping platforms*. *2023*, 1–10. https://doi.org/10.1057/s41599-023-02277-7

Liang, T. (2022). *A study on the influence of online reviews of new products on consumers ' purchase decisions : An empirical study on*. *September*, 1–22. https://doi.org/10.3389/fpsyg.2022.983060

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. *1*. http://arxiv.org/abs/1907.11692

Mukherjee, R. K. (2024). Transforming Digitally: A Systematic Literature Review of Consumer Behavior & Digital Transformation Interaction. *International Journal of Development Research*, *14*(2015), 67052–67062. https://doi.org/doi.org/10.37118/ijdr.28942.11.2024

Opitz, J. (2024). *A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice*. *12*(2018), 820–836.

Purnama, A. (2023). Analisis Sentimen Terhadap Review Pengguna Indrive di Google Playstore Menggunakan Algoritma Support Vector Machine. *Jurnal Darma Agung*, *31*(1), 1015–1025.

Purwati, K. Y. (2023). Analisis Sentimen Berita Vaksin COVID-19 dengan Robustly Optimized BERT Pre-Training Approach (RoBERTa). *Skripsi, UIN Syarif Hidayatullah Jakarta*.

Roberta, M. (n.d.). *Prediksi Sentimen Pada Teks Media Sosial Corporate University*. 302–316.

Sandra, L., Marcel, Gunarso, G., Fredicia, & Riruma, O. W. (2022). Are University Students Independent: Twitter Sentiment Analysis of Independent Learning in Independent Campus Using RoBERTa Base IndoLEM Sentiment Classifier Model. *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, 249–253. https://doi.org/10.1109/ISMODE53584.2022.9743110

Sukma, F. D., Wijaya, R., & Romadhony, A. (2025). Speech to Text Correction for Indonesian Early Marriage Counseling Chatbots Using. *Journal on Computing*, *10*(August), 72–87. https://doi.org/10.21108/indojc.v10i1.9708

Walther, M., Watson, S., Boden, A., Stel, M., Walther, M., Watson, S., Boden, A., A, M. S., & Walther, M. (2026). *A grounded theory of how consumers determine the veracity of online user reviews*. *3001*. https://doi.org/10.1080/0144929X.2025.2528764