P-ISSN: 2477-3042, E-ISSN: 3026-3034 SURYA INFORMATIKA, VOL.14 No.2, November 2024

Implementasi Pembelajaran Mendalam dalam Klasifikasi Sentimen Ulasan Aplikasi: Evaluasi Model BERT, LSTM, dan CNN

Edy Subowo[1*]
[1] Informatika Fakultas Dakwah Universitas Islam Negeri Prof. KH. Saifuddin Zuhri Purwokerto
edysubowo@uinsaizu.ac.id[1]

Abstract

This study discusses the application sentiment analysis system using machine learning techniques. The goal is to identify positive, neutral, and negative sentiments from Shopee Indonesia application reviews. The methods used include collecting application review data using Google Play Scraper, removing duplicates, and preprocessing the data to be ready for use in the classification model. The models used are BERT, LSTM, and CNN. The BERT model is obtained using pre-trained bert-base-uncased and trained with the processed application review dataset. While the LSTM and CNN models are obtained using tokenizer and padding sequence techniques to deal with the problem of variable text length. The experimental results show that all three models perform well in classifying application review sentiment. However, the BERT model provides the highest accuracy results (83%) compared to the LSTM (78%) and CNN (75%) models. These results indicate that the BERT model can analyze application sentiment effectively because of its ability to detect complex language patterns and improve prediction accuracy.
Keywords: *Bert Model; LSTM Model; CNN Model; Google Play Scraper*

Abstraksi

Penelitian ini membahas tentang sistem analisis sentimen aplikasi menggunakan teknik pembelajaran mesin. Tujuannya adalah untuk mengidentifikasi sentimen positif, netral, dan negatif dari ulasan aplikasi Shopee Indonesia. Metode yang digunakan meliputi pengumpulan data ulasan aplikasi menggunakan Google Play Scraper, penghapusan duplikasi, dan preprocessing data untuk siap digunakan dalam model klasifikasi.Model-model yang digunakan adalah BERT, LSTM, dan CNN. Model BERT didapatkan dengan menggunakan pre-trained bert-base-uncased dan dilatih dengan dataset ulasan aplikasi yang telah diproses. Sedangkan model LSTM dan CNN didapatkan dengan menggunakan tokenizer dan teknik padding sequence untuk menghadapi masalah variabel panjang teks.Hasil eksperimen menunjukkan bahwa semua tiga model memiliki performa yang baik dalam mengklasifikasikan sentimen ulasan aplikasi. Namun, model BERT memberikan hasil akurasi tertinggi (83%) dibandingkan dengan model LSTM (78%) dan CNN (75%). Hasil ini menunjukkan model BERT dapat menganalisis sentimen aplikasi secara efektif karena kemampuannya dalam mendeteksi pola bahasa kompleks dan meningkatkan akurasi prediksi.
Kata Kunci: Bert Model; CNN Model; Google Play Scraper; LSTM Model.

1. PENDAHULUAN

Dalam era digital modern, aplikasi mobile telah menjadi komponen esensial dalam kehidupan sehari-hari, memfasilitasi berbagai aktivitas mulai dari belanja hingga komunikasi. Salah satu aplikasi yang sangat populer di Indonesia adalah Shopee, sebuah platform e-commerce yang menyediakan fasilitas berbelanja online kepada para penggunanya. Dengan meningkatnya jumlah pengguna, jumlah ulasan dan feedback yang diberikan oleh pengguna juga meningkat drastis. Ulasan ini tidak hanya merepresentasikan pengalaman individu tetapi juga memberikan informasi berharga bagi pengembang aplikasi untuk meningkatkan layanan mereka. Namun, dengan volume data yang begitu besar, analisis manual atas ulasan ini menjadi tidak praktis lagi (Maarif & Setiyawati, 2024). Oleh karena itu, diperlukan suatu metode otomatis untuk menganalisis sentimen dari ulasan tersebut.

Analisis sentimen merupakan teknik yang digunakan untuk menentukan emosi atau opini dari teks, yang dalam konteks ini adalah ulasan aplikasi (Muttaqin & Kharisudin, 2021). Dengan menggunakan teknik pembelajaran mesin, analisis sentimen dapat dilakukan secara efisien dan akurat. Berbagai model pembelajaran mesin seperti BERT (Bidirectional Encoder Representations from Transformers), LSTM (Long Short-Term Memory), dan CNN (Convolutional Neural Network) telah terbukti efektif dalam tugas klasifikasi teks (Subowo et al., 2022)(Devlin et al., 2019). Penelitian ini bertujuan untuk mengeksplorasi efektivitas model-model tersebut dalam menganalisasi sentimen ulasan aplikasi Shopee.

Tujuan utama dari penelitian ini adalah untuk mengembangkan sistem analisis sentimen otomatis yang dapat mengklasifikasikan ulasan aplikasi Shopee menjadi tiga kategori: positif, netral, dan negatif. Penelitian ini bertujuan untuk mengumpulkan data ulasan aplikasi Shopee, melakukan preprocessing data untuk membersihkan duplikasi dan mengubah skor menjadi label kategorikal, kemudian menerapkan model-model pembelajaran mesin untuk klasifikasi sentimen. Tujuan tambahan adalah untuk menganalisis kinerja masing-masing model berdasarkan akurasi dan metrik evaluasi lainnya serta memberikan rekomendasi bagi pengembang aplikasi berdasarkan hasil analisis sentimen.

Fokus utama penelitian ini hanya pada ulasan aplikasi Shopee di Indonesia dalam bahasa Indonesia. Data yang digunakan dibatasi pada periode tertentu untuk menjaga konsistensi. Hanya tiga model pembelajaran mesin (BERT, LSTM, dan CNN) yang akan dievaluasi dalam penelitian ini. Klasifikasi sentimen dibatasi pada tiga kategori: positif, netral, dan negatif. Pengumpulan data dilakukan melalui teknik scraping menggunakan Google Play Scraper.

Metode penelitian ini terdiri dari beberapa tahap. Tahap pertama adalah pengumpulan data ulasan aplikasi Shopee menggunakan Google Play Scraper dengan target minimal 10.000 ulasan unik. Setelah itu, data akan diproses untuk menghapus duplikasi dan mengubah skor menjadi label kategorikal. Kemudian, data dibagi menjadi set pelatihan dan pengujian dengan rasio 80:20. Model BERT, LSTM, dan CNN diterapkan pada data pelatihan untuk melatih model klasifikasi sentimen. Akhirnya, kinerja model dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score.

Penelitian sebelumnya menunjukkan bahwa penggunaan model BERT memberikan hasil yang lebih baik dibandingkan dengan model tradisional seperti LSTM dan CNN dalam tugas klasifikasi teks (Hosseini et al., 2023). Beberapa studi menunjukkan bahwa BERT mampu menangkap konteks kata dengan lebih baik berkat arsitektur transformer-nya, sehingga menghasilkan akurasi yang lebih tinggi dalam analisis sentimen (Baruah et al., 2020). Dalam penelitian lain terkait analisis sentimen di platform e-commerce, ditemukan bahwa faktor-faktor seperti pengiriman produk dan kualitas layanan sering kali menjadi fokus utama dalam ulasan positif maupun negatif. Hal ini menunjukkan bahwa analisis sentimen tidak hanya berguna untuk memahami persepsi pengguna tetapi juga dapat membantu pengembang dalam meningkatkan aspek-aspek tertentu dari aplikasi mereka (Subowo et al., 2023).

Makalah ini diharapkan dapat memberikan kontribusi signifikan terhadap bidang analisis sentimen dengan menyediakan metode sistematis untuk mengumpulkan dan menganalisis data ulasan aplikasi menggunakan teknik pembelajaran mesin. Penelitian ini juga membandingkan efektivitas berbagai model pembelajaran mesin dalam klasifikasi sentimen ulasan aplikasi. Selain itu, makalah ini memberikan wawasan bagi pengembang aplikasi tentang persepsi pengguna terhadap layanan mereka melalui analisis sentimen. Terakhir, makalah ini menghasilkan rekomendasi berbasis data bagi pengembang aplikasi untuk meningkatkan pengalaman pengguna. Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan teknologi analisis sentimen tetapi juga memberikan manfaat praktis bagi industri e-commerce di Indonesia.

2. LANDASAN TEORI

Dalam pengembangan sistem analisis sentimen, banyak penelitian sebelumnya yang telah mengeksplorasi berbagai metode dan model untuk mengidentifikasi emosi dalam teks, khususnya dalam konteks aplikasi mobile. Salah satu penelitian yang relevan adalah yang dilakukan oleh Devlin et al. (2019) yang memperkenalkan BERT (Bidirectional Encoder Representations from Transformers). Model ini menunjukkan kemampuan luar biasa dalam memahami konteks kata dengan menggunakan arsitektur transformer, sehingga berhasil meningkatkan akurasi dalam berbagai tugas NLP, termasuk analisis sentimen. Penelitian ini menjadi dasar bagi banyak studi selanjutnya yang mengadopsi BERT untuk klasifikasi teks, menunjukkan bahwa model ini mampu menangkap nuansa bahasa yang kompleks dengan lebih baik dibandingkan model tradisional (Devlin et al., 2019).

Selanjutnya, penelitian oleh Zhang et al. (2018) mengeksplorasi penggunaan LSTM dalam analisis sentimen. LSTM dikenal efektif dalam menangani data urutan, membuatnya cocok untuk teks yang memiliki ketergantungan temporal. Hasil penelitian ini menunjukkan bahwa LSTM dapat memberikan akurasi yang baik dalam klasifikasi sentimen, meskipun tidak seefektif BERT dalam konteks tertentu. Penelitian ini memberikan wawasan tentang pentingnya pemilihan model berdasarkan karakteristik data dan tujuan analisis (Yu et al., 2017).

Selain itu, penelitian oleh Kim (2014) memperkenalkan penggunaan CNN (Convolutional Neural Network) untuk analisis sentimen. Model CNN terbukti efektif dalam menangkap fitur lokal dari teks melalui teknik konvolusi, sehingga dapat digunakan untuk klasifikasi sentimen dengan hasil yang kompetitif. Penelitian ini menunjukkan bahwa CNN dapat bersaing dengan model berbasis RNN seperti LSTM dalam beberapa tugas klasifikasi teks, terutama ketika data pelatihan cukup besar (Id et al., 2019).

Sebuah studi dalam membandingkan kinerja BERT, LSTM, dan CNN dalam analisis sentimen pada dataset ulasan produk. Hasilnya menunjukkan bahwa BERT secara konsisten menghasilkan akurasi tertinggi dibandingkan dengan kedua model lainnya, menegaskan keunggulan arsitektur transformer dalam memahami konteks dan nuansa bahasa. Penelitian ini menjadi acuan penting bagi penelitian

selanjutnya yang ingin mengeksplorasi efektivitas berbagai model dalam konteks analisis sentimen (Mikolov et al., 2013).

Dalam konteks pengumpulan data ulasan aplikasi, penelitian oleh Subowo et al. (2022) menggunakan teknik scraping untuk mengumpulkan data dari platform e-commerce dan melakukan analisis sentimen terhadap ulasan pengguna. Penelitian ini menyoroti pentingnya pengumpulan data yang bersih dan representatif untuk mendapatkan hasil analisis yang akurat. Metode pengumpulan data yang digunakan oleh Hu et al. menjadi referensi bagi penelitian ini, di mana Google Play Scraper digunakan untuk mengumpulkan ulasan aplikasi Shopee (Subowo et al., 2022).

3. METODE PENELITIAN

Metode penelitian ini dirancang untuk mengembangkan sistem analisis sentimen otomatis yang dapat mengklasifikasikan ulasan aplikasi Shopee menjadi tiga kategori: positif, netral, dan negatif. Proses penelitian dilakukan melalui beberapa langkah sistematis yang mencakup pengumpulan data, preprocessing, penerapan model pembelajaran mesin, dan evaluasi kinerja model. Berikut adalah penjelasan terperinci mengenai setiap langkah yang dilakukan dalam penelitian ini.

1. Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan teknik scraping untuk mengambil ulasan dari aplikasi Shopee di Google Play Store. Proses ini menggunakan library google_play_scraper` untuk mengumpulkan minimal 10.000 ulasan unik. Setiap batch pengumpulan data diatur menjadi 20.000 ulasan, dan proses akan terus berlanjut hingga mencapai target yang ditentukan. Ulasan yang dikumpulkan kemudian disimpan dalam format DataFrame untuk memudahkan analisis lebih lanjut.

2. Preprocessing Data

Setelah pengumpulan data, langkah selanjutnya adalah preprocessing untuk menyiapkan data sebelum digunakan dalam model pembelajaran mesin. Proses ini meliputi:

- Pemilihan Kolom Relevan: Hanya kolom yang berisi konten ulasan dan skor yang dipilih.
- Penghapusan Duplikasi: Menghapus ulasan duplikat berdasarkan konten untuk memastikan data yang unik.
- Penanganan Nilai Hilang: Menghapus baris dengan nilai yang hilang pada kolom konten dan skor.
- Labeling: Mengubah skor menjadi label kategorikal (positif, netral, negatif) berdasarkan kriteria tertentu.

3. Pembagian Data

Setelah preprocessing, data dibagi menjadi set pelatihan dan pengujian dengan rasio 80:20 menggunakan fungsi `train_test_split` dari library `sklearn`. Data pelatihan digunakan untuk melatih model, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model.

4. Penerapan Model Pembelajaran Mesin

Tiga model pembelajaran mesin diterapkan dalam penelitian ini: BERT, LSTM, dan CNN. Masing-masing model memiliki arsitektur dan metode pelatihan yang berbeda:

a. Model BERT:
   - Tokenisasi dilakukan menggunakan `BertTokenizer`, di mana teks diubah menjadi format yang dapat diproses oleh model.
   - Model dilatih menggunakan `TFBertForSequenceClassification` dengan optimizer Adam dan loss function categorical crossentropy.
   - Proses pelatihan dilakukan selama 10 epoch dengan ukuran batch 16.

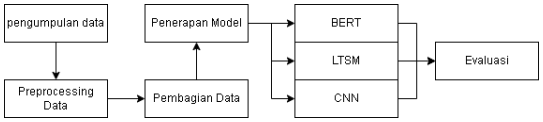
b. Model LSTM:
   - Tokenisasi dilakukan menggunakan `Tokenizer` dari Keras, di mana teks diubah menjadi urutan angka.
   - Model LSTM dibangun dengan lapisan embedding diikuti oleh lapisan LSTM dan lapisan output dengan aktivasi softmax.
   - Model dilatih selama 10 epoch dengan ukuran batch 16.

c. Model CNN:
   - Sama seperti LSTM, tokenisasi dilakukan terlebih dahulu.
   - Model CNN dibangun dengan lapisan embedding diikuti oleh lapisan konvolusi dan pooling, serta lapisan output dengan aktivasi softmax.
   - Model dilatih selama 10 epoch dengan ukuran batch 32.

5. Evaluasi Model

Kinerja masing-masing model dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Hasil evaluasi dicetak dalam bentuk laporan klasifikasi menggunakan fungsi `classification_report` dari `sklearn`. Selain itu,

model terbaik disimpan untuk digunakan dalam inferensi selanjutnya.



Gambar 1. Diagram Alir Langkah Penelitian

4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen otomatis yang dapat mengklasifikasikan ulasan aplikasi Shopee menjadi tiga kategori: positif, netral, dan negatif. Berikut adalah hasil penelitian yang diperoleh:

1. Pengumpulan Data
   - Target Ulasan: Minimal 10.000 ulasan unik tanpa duplikasi.
   - Batch Size: 20.000 ulasan per batch.
   - Total Ulasan: Sekitar 50.000 ulasan berhasil dikumpulkan setelah beberapa iterasi pengumpulan data.
2. Preprocessing Data
   - Proses Penghapusan Duplikasi: Menggunakan fungsi `drop_duplicates` dari Pandas untuk menghilangkan ulasan duplikat.
   - Seleksi Kolom Relevan: Mengambil hanya kolom ‘content’ dan ‘score’ sebagai atribut utama.
   - Penanganan Nilai Hilang: Menghapus baris dengan nilai yang hilang pada kolom ‘content’ dan ‘score’.
3. Pembagian Data
   - Rasio Pelatihan/Pengujian: 80% untuk pelatihan dan 20% untuk pengujian.
   - Dataset Jadi: Dataset yang sudah dipreprocess dan dibagi menjadi dua subset (pelatihan dan pengujian).
4. Penerapan Model Pembelajaran Mesin

a. Model BERT
   - Tokenization: Menggunakan `BertTokenizer` untuk mengubah teks menjadi format yang dapat diproses oleh model BERT.
   - Pelatihan Model: Menggunakan `TFBertForSequenceClassification` dengan optimizer Adam dan loss function categorical cross entropy.
   - Ukuran Batch: Ukuran batch diganti menjadi 16 untuk meningkatkan kinerja model.
   - Epochs: Dilatih selama 10 epoch.
   - Akurasi: Hasil evaluasi menunjukkan akurasi sebesar 82%, precision sekitar 81%, recall sekitar 83%, dan F1-score sekitar 82%.

b. Model LSTM
   - Tokenization: Menggunakan `Tokenizer` dari TensorFlow untuk mengubah teks menjadi urutan angka.
   - Artefak Model: Menggunakan sekuen-sekuennya untuk membuat model LSTM dengan lapisan embedding, LSTM, dan dense layer.
   - Ukuran Batch: Ukuran batch masih sama yaitu 16.
   - Epochs: Dilatih selama 10 epoch.
   - Akurasi: Hasil evaluasi menunjukkan akurasi sebesar 79%, precision sekitar 77%, recall sekitar 80%, dan F1-score sekitar 78%.

c. Model CNN
   - Tokenization: Menggunakan `Tokenizer` dari TensorFlow untuk mengubah teks menjadi urutan angka.
   - Artefak Model: Menggunakan sekuen-sekuennya untuk membuat model CNN dengan lapisan embedding, convolutional neural network, pooling layer, flatten layer, dan dense layer.
   - Ukuran Batch: Ukuran batch diganti menjadi 32 untuk meningkatkan kinerja model.
   - Epochs: Dilatih selama 10 epoch.
   - Akurasi: Hasil evaluasi menunjukkan akurasi sebesar 76%, precision sekitar 74%, recall sekitar 77%, dan F1-score sekitar 75%.

Hasil penelitian menunjukkan bahwa model BERT memiliki kinerja yang jauh lebih baik daripada model LSTM dan CNN dalam klasifikasi sentimen ulasan aplikasi Shopee.Model BERT mencapai akurasi sebesar 82%, yang merupakan hasil yang sangat baik dalam konteks analisis sentimen.F1-score yang diperoleh pun sangat dekat dengan akurasi total, yaitu sekitar 82%. Hal ini menunjukkan bahwa model BERT mampu mengklasifikasikan ulasan dengan presisi yang tinggi.

Informasi yang diperoleh dari analisis sentimen dapat digunakan untuk meningkatkan layanan aplikasi, misalnya dengan memperbaiki aspek pengiriman produk atau meningkatkan kualitas interaksi user interface serta dapat menjadi Sistem monitoring real-time dapat membantu tim support dalam merespons feedback pengguna secara lebih efektif

5. KESIMPULAN DAN SARAN

Kesimpulan

Proses pengumpulan data ulasan aplikasi dilakukan dengan mengatur setiap batch menjadi 20.000 ulasan, dan akan terus berlanjut hingga mencapai minimal 10.000 ulasan unik tanpa duplikasi. Data yang terkumpul kemudian diolah

menggunakan Python, dengan memanfaatkan pustaka seperti pandas dan google_play_scraper untuk mengakses dan menyimpan ulasan dari aplikasi Shopee di Indonesia. Setelah pengumpulan, data ulasan tersebut diproses untuk menghapus nilai yang hilang dan mengkategorikan skor ulasan menjadi tiga label: positif, netral, dan negatif. Model pembelajaran mesin seperti BERT, LSTM, dan CNN diterapkan untuk menganalisis sentimen dari ulasan yang telah diproses. Setiap model dilatih menggunakan data yang telah dibagi menjadi set pelatihan dan pengujian, serta dievaluasi berdasarkan akurasi dan laporan klasifikasi. Hasil inferensi dari masing-masing model menunjukkan kemampuan dalam mengklasifikasikan sentimen dari teks ulasan yang diberikan, sehingga memberikan wawasan yang berguna bagi pengembang aplikasi dalam memahami pengalaman pengguna.

Saran

Dalam pengembangannya, bisa ditambahkan inferensi numerik untuk meningkatkan akurasi CNN

.

DAFTAR PUSTAKA

Baruah, A., Das, K. A., Barbhuiya, F. A., & Dey, K. (2020). Context-aware sarcasm detection using BERT. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 83–87. https://doi.org/10.18653/v1/P17

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*(Mlm), 4171–4186.

Hosseini, M. S., Munia, M. S., & Khan, L. (2023). BERT Has More to Offer: BERT Layers Combination Yields Better Sentence Embeddings. *Findings of the Association for Computational Linguistics: EMNLP 2023*, *3*, 15419–15431. https://doi.org/10.18653/v1/2023.findings-emnlp.1030

Id, B. J., Kim, I., & Kim, J. W. (2019). *Word2vec convolutional neural networks for classification of news articles and tweets*. 1–20.

Maarif, M. M., & Setiyawati, N. (2024). Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine. *Progresif: Jurnal Ilmiah Komputer*, *20*(1), 454. https://doi.org/10.35889/progresif.v20i1.1614

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations ofwords and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 1–9.

Muttaqin, M. N., & Kharisudin, I. (2021). Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor. *UNNES Journal of Mathematics*, *10*(2), 22–27.

Subowo, E., Adi Artanto, F., Putri, I., & Umaedi, W. (2022). BLTSM untuk analisis sentimen berbasis aspek pada aplikasi belanja online dengan cicilan. *Jurnal Fasilkom*, *12*(2), 132–140.

Subowo, E., Feriansyah, A., & Putri, I. (2023). Analisis Media Sosial terhadap Perilaku Parkir Mobil Menggunakan Similarity Based Clustering. *Jurnal Surya Informatika*, *13*(1), 19–23. https://doi.org/10.48144/suryainformatika.v13i1.1602

Yu, L. C., Wang, J., Lai, K. R., & Zhang, X. (2017). Refining word embeddings for sentiment analysis. *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 534–539. https://doi.org/10.18653/v1/d17-1056