
SISTEM ANALISIS SENTIMEN PRODUK PADA APLIKASI LAZADA MENGGUNAKAN METODE NAIVE BAYES

Alvina Gusti Pramita¹, Fajar Nugraha²

Universitas Muria Kudus, Sistem Informasi^{1,2}

Jl. Lingkar Utara UMK, Gondangmanis, Bae, Kudus

e-mail: 202153018@std.umk.ac.id¹, fajar.nugraha@umk.ac.id²

Abstrak

Lazada merupakan salah satu e-commerce yang sangat populer di Indonesia. Sebagai platform e-commerce terkemuka, Lazada menghadapi berbagai tantangan besar dalam mengelola review dari pengguna. Dengan adanya jutaan bahkan ribuan review yang beragam, mengidentifikasi kualitas produk sehingga dapat memenuhi harapan pelanggan bisa menjadi tugas yang sangat rumit. Selama ini, perusahaan melakukan analisa review produk Lazada yang sudah dipasarkan dengan menghabiskan banyak waktu dan sumber daya manusia. Untuk menganalisis sentimen dari review pelanggan, proses pengumpulan data sampai dengan hasil sentimen dilakukan secara manual. Selain itu, ketersediaan data yang banyak akan sulit untuk secara efisien mengolah dan menganalisis semua review secara manual. Maka dari itu, diperlukan suatu sistem yang mampu mengatasi permasalahan yang terjadi pada proses analisis ulasan secara manual. Di dunia kecerdasan buatan, permasalahan seperti ini termasuk ke dalam domain Natural Language Processing (NLP). Dengan adanya penelitian ini akan membantu proses review ulasan dari pengguna Lazada dengan menggunakan aplikasi SMARTSHOP-REVIEW dengan model Naïve Bayes, mengingat bahwa pola sentiment yang ada dari ulasan pengguna itu sangat bermacam-macam. Data yang digunakan adalah data dari Lazada Indonesian Review 2019 yang berisi 38.071 review. Dari data tersebut, akan dikelompokkan dari sentimen yang ada yaitu Positif dan Negatif dengan hasil Positif sebesar 32.198 dan Negatif sebesar 5.753.

Kata kunci: Analisa Sentimen, Naïve Bayes, NLP, Ulasan Produk, Smartshop

Abstract

Lazada is one of the most popular e-commerce sites in Indonesia. As a leading e-commerce platform, Lazada faces major challenges in managing user reviews. With millions or even thousands of varying reviews, identifying product quality so that it meets customer expectations can be a very complicated task. So far, the company has been analyzing reviews of Lazada products that have been marketed using a lot of time and human resources. To analyze sentiment from customer reviews, the data collection process to arrive at sentiment results is carried out manually. In addition, the availability of large amounts of data makes it difficult to efficiently process and analyze all reviews manually. Therefore, we need a system that is able to overcome problems that occur in the manual review analysis process. In the world of artificial intelligence, problems like this fall into the domain of Natural Language Processing (NLP). This research will help the process of reviewing reviews from Lazada users by using the SMARTSHOP-REVIEW application with the Naïve Bayes model, remembering that the sentiment patterns that exist from user reviews vary greatly. The data used is data from the 2019 Lazada Indonesian Review which contains 38,071 reviews. From this data, existing sentiment will be grouped into Positive and Negative with Positive results of 32,198 and Negative of 5,753.

Keywords: Sentiment Analysis, Naïve Bayes, NLP, Product Review, Smartshop

1. PENDAHULUAN

Lazada sebagai platform e-commerce terkemuka menghadapi beberapa tantangan besar dalam mengelola review pengguna [1]. Pertama, mereka merasa sulit mengevaluasi kualitas produk berdasarkan tinjauan yang banyak dan beragam. Dengan jutaan review yang beragam, mengidentifikasi kualitas produk sehingga dapat memenuhi harapan pelanggan bisa menjadi tugas yang sangat rumit. Selain itu, memoderasi review yang tidak pantas atau menyesatkan juga merupakan masalah yang signifikan. Selama ini, perusahaan melakukan menganalisis review produk Lazada yang sudah dipasarkan dengan menghabiskan banyak waktu dan sumber daya manusia. Untuk menganalisis sentimen dari review pelanggan, proses pengumpulan data sampai dengan hasil sentimen dilakukan dengan manual. selain itu, ketersediaan data yang banyak akan sulit untuk secara efisien mengolah dan menganalisis semuanya secara manual [2]. Dari data Lazada Indonesian Review 2019, terdapat data review content sebanyak 288.515. Dengan banyaknya review yang ada, proses analisis review dari pelanggan sangat memakan waktu jika dilakukan secara manual.

Maka dari itu, diperlukan suatu sistem yang baru untuk mengatasi permasalahan yang terjadi pada proses analisis review secara manual. Salah satu solusi yang dapat dilakukan adalah dengan membuat suatu sistem analisis review yang terotomatisasi. Proses analisis review dapat dilakukan oleh komputer yang terlebih dahulu diprogram untuk mengenali serta memahami bahasa manusia. Di dunia kecerdasan buatan, permasalahan seperti ini termasuk ke dalam domain Natural Language Processing (NLP) [3]. Dengan solusi yang kami kembangkan ini, analisis sentimen yang akan diimplementasikan membuat proses analisis tinjauan menjadi lebih efisien. Proses ekstraksi data sampai dengan analisis sentimen akan dilakukan secara otomatis. Solusi ini akan mengurangi waktu dan sumber daya yang dibutuhkan serta meningkatkan kemampuan untuk merespons perubahan sentimen pelanggan dengan lebih cepat dan lebih tepat [4].

Proyek sentimen analisis pada aplikasi Lazada ini bertujuan untuk menggali pandangan dan penilaian pelanggan terhadap berbagai produk yang terdaftar di platform Lazada. Penilaian atau sentimen mengacu pada suatu topik tertentu [5] [6]. Pernyataan pada topik yang sama mungkin memiliki makna yang berbeda dengan pernyataan yang sama pada topik yang berbeda. Oleh karena itu, beberapa penelitian perlu mengidentifikasi topik yang sedang dibahas sebelum melakukan analisis sentimen [7]. Dengan menggunakan teknik analisis sentimen pada ulasan-ulasan pelanggan, proyek ini akan menghasilkan sebuah dashboard atau website interaktif yang dapat memvisualisasikan hasil analisis sentimen, memberikan wawasan tentang sejauh mana pelanggan merasa puas atau tidak puas dengan produk tertentu, serta mengidentifikasi masalah atau area perbaikan potensial. Web aplikasi ini akan memberikan manfaat yang berharga bagi perusahaan untuk mengambil keputusan yang lebih baik dalam pengelolaan produk, pemasaran, dan pengembangan produk.

2. METODE PENELITIAN

Ada dua metode yang digunakan pada pengembangan sistem aplikasi Smartshop-Review ini, yaitu:

2.1 Metode Pengumpulan Data

Proses awal yang dilakukan adalah mengambil dan mengamati data ulasan produk pada aplikasi Lazada dari Kaggle. Data ini dianalisis untuk mengetahui sentimen pengguna terhadap produk yang direview. Kemudian dilakukan pencarian informasi dan teori yang relevan dengan penelitian ini

2.2 Metode Pengembangan Sistem

Metode pengembangan sistem adalah merupakan sebuah teknik untuk membangun atau mengubah sistem dengan memanfaatkan model-model yang telah terbukti efektif [8]. Dalam membangun sistem pada penelitian ini menggunakan *tools Steamlit* dengan menggunakan *Google Collab*.

3. HASIL DAN PEMBAHASAN

Proyek ini berhasil mengembangkan aplikasi Smartshop Review dengan fokus pada analisis sentimen ulasan produk di platform e-commerce Lazada menggunakan algoritma Naive Bayes. Berikut adalah hasil yang diperoleh dari berbagai tahapan proyek ini:

A. Problem Scoping

Proyek Akhir ini berfokus pada Sentiment Analysis ulasan produk di platform e-commerce Lazada. Sentiment Analysis sendiri merupakan teknik penting dalam pemrosesan bahasa alami yang memungkinkan kita untuk memahami dan mengklasifikasikan perasaan yang terkandung dalam teks, salah satunya ulasan produk. Dengan menggunakan algoritma Naive Bayes, proyek ini bertujuan untuk mengembangkan model yang dapat mengklasifikasikan ulasan produk ke dalam dua kategori sentimen: positif dan negatif.

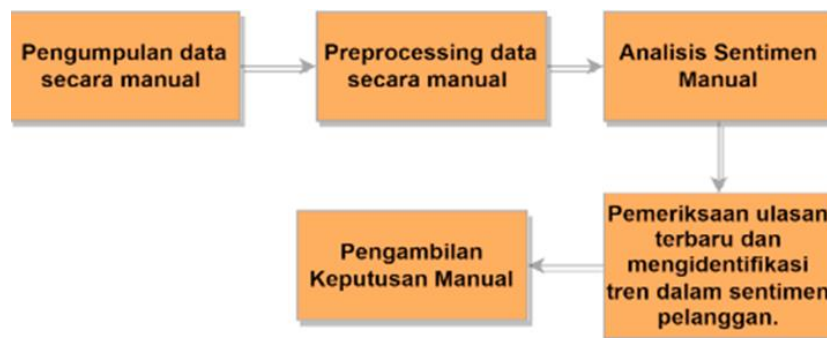
Proyek ini memiliki tujuan membantu penjual memahami ulasan produk mereka, membantu konsumen dapat lebih mudah menemukan produk yang sesuai dengan preferensi mereka berdasarkan ulasan produk dari pengguna laidan mengembangkan model Sentiment Analysis yang dapat memberikan wawasan berharga untuk kemajuan platform e-commerce Lazada secara keseluruhan [9].

Ekspektasi yang diinginkan oleh Business User dalam proyek Sentiment Analysis pada aplikasi Lazada diantaranya adalah pengukuran yang akurat terhadap kepuasan pelanggan, pendidentifikasi masalah dengan cepat, serta perbaikan produk dan layanan. Tentu saja ekspektasi tersebut akan berujung pada pengambilan keputusan strategis yang lebih baik. Perusahaan mengharapkan hasil analisis sentimen dapat memberikan wawasan yang berharga untuk meningkatkan pengalaman pelanggan atau customer experience.

Terdapat batasan-batasan yang dipertimbangkan oleh Business User dalam proyek analisis sentimen pada aplikasi Lazada adalah:

- 1) Biaya: Pengembangan dan pelaksanaan proyek Sentiment Analysis dapat memerlukan anggaran yang signifikan untuk sumber daya komputasi, alat, atau personel yang dibutuhkan.
- 2) Tantangan dalam Bahasa atau Wilayah Tertentu: Bahasa atau wilayah yang berbeda dapat menimbulkan tantangan karena model analisis sentimen harus diperluas atau disesuaikan untuk masing-masing bahasa daerah atau wilayah.
- 3) Kualitas Data: Kualitas data yang tidak memadai atau data yang tidak terstruktur dapat menjadi batasan. Business User harus memastikan bahwa data yang digunakan untuk analisis memiliki kualitas yang memadai.
- 4) Volume Data: Jika volume data sangat besar, pengolahan dan analisis data dapat menjadi lambat dan memerlukan sumber daya komputasi yang besar.

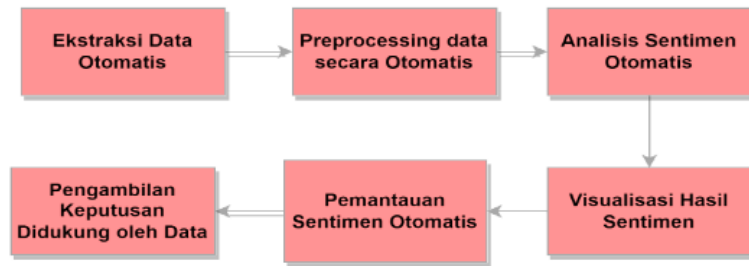
Dengan melakukan analisis sentimen produk Lazada, perusahaan dapat meningkatkan kualitas produk yang dipasarkan serta dapat meningkatkan penjualan di platform tersebut. Solusi ini cocok untuk menyelesaikan masalah bisnis yang berkaitan dengan customer experience.



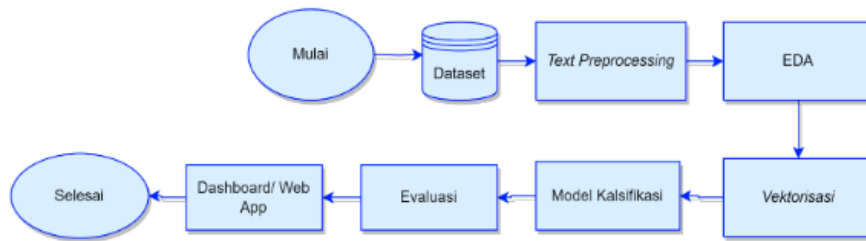
Gambar 1. Proses analisis review pelanggan yang secara manual (As-Is)

Gambar 1. adalah proses analisis review yang saat ini dilakukan (As-Is). Saat ini, perusahaan melakukan analisis sentimen secara manual. Berawal dari dikumpulkannya data ulasan pelanggan secara manual dan kemudian dilanjutkan dengan mengolahnya menjadi data terstruktur. Ulasan pelanggan yang sudah menjadi data berbentuk tabel tersebut kemudian dianalisis secara manual satu persatu. Keluaran analisis manual tersebut berupa wawan atau insight yang akan digunakan untuk penentuan suatu keputusan bisnis.

Proses analisis review pelanggan yang sedang dilakukan secara manual dapat berubah menjadi (To-Be) seperti pada Gambar 2. Pengumpulan data beserta tahap pre-processing dilakukan secara otomatis dan berkala. Data ulasan yang sudah diproses tersebut kemudian dianalisis secara otomatis dengan model yang telah dibuat dan dirancang. Keluaran dari analisis tersebut berupa sentimen yang divisualisasikan pada dashboard sehingga business user dapat melihat secara langsung dan segera menentukan suatu keputusan bisnis.



Gambar 2. Proses analisis review pelanggan secara otomatis (To-Be)



Gambar 3. Langkah-langkah dalam analisis sentimen prosuk Lazada

Langkah-langkah dalam analisis sentimen produk Lazada dimulai dengan pengumpulan data ulasan pelanggan dari platform Lazada. Pengumpulan data tersebut diantaranya adalah deskripsi produk, harga, dan peringkat. Selanjutnya, dilakukan preprocessing data untuk membersihkan teks seperti menghilangkan tanda baca serta mengubah teks ke dalam bentuk yang dapat dianalisis. Kemudian, dilakukan exploratory data analysis (EDA) untuk memahami data ulasan pelanggan. Tahapan berikutnya dilanjutkan dengan klasifikasi ulasan ke dalam kategori sentimen positif, negatif, atau netral menggunakan salah satu model pemrosesan bahasa alami (NLP) yaitu model Naive Bayes. Setelah model analisis sentimen dibuat dan dilatih, evaluasi model dilakukan untuk memastikan akurasi dan kinerja yang baik. Terakhir, hasil analisis disajikan melalui dashboard atau aplikasi web yang dapat memvisualisasikan informasi sentimen secara interaktif untuk memberikan pemahaman yang lebih baik kepada pengguna bisnis. Solusi ini memberikan business user alat yang kuat untuk mengambil keputusan berdasarkan data yang berkualitas dan merespons perubahan pasar dengan lebih cepat dan lebih tepat. Business user dapat mengakses dashboard yang memvisualisasikan sentimen produk secara real-time. Dengan demikian, solusi ini dapat meningkatkan efisiensi operasional dan kemampuan bisnis untuk tetap bersaing di pasar e-commerce seperti Lazada..

B. Data Acquisition

Data yang digunakan dalam proyek analisis sentimen bersumber dari platform Kaggle yaitu data product review pada platform Lazada yang terdiri dari dua data yaitu data items dan data review. Kedua data tersebut kemudian digabungkan menggunakan teknik *left join* dengan “*itemid*” sebagai kolom kuncinya. Penggabungan kedua data menghasil sebanyak 38071 review dari product review Lazada.

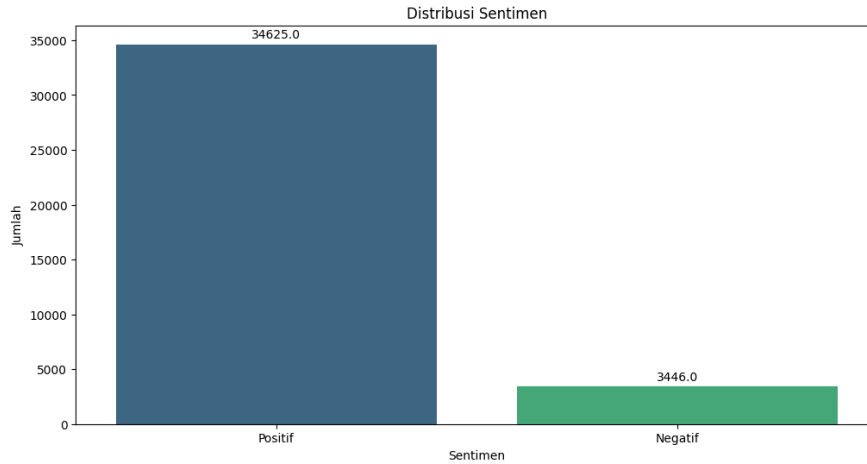
C. Data Exploration

Pada data yang kami gunakan yaitu data review Lazada yang berisi 38.071 review ini kita lakukan eksplorasi dengan membuat dan membagi sentiment yang ada pada setiap review nya. Semua data digabungkan menjadi dataset gabung dan dibuat kamus penggantian sentiment mapping yaitu Positif dengan nilai 1 dan Negatif dengan nilai 2. Selanjutnya dibuat sentiment count dari dataset gabung diatas sebagai berikut:

```
sentimen_count = dt_gabungan['Sentimen'].value_counts()
```

```
sentimen_count  
Hasilnya :  
1      32198  
0      5753  
Name: Sentimen, dtype: int64
```

Dari hasil diatas didapatkan review dengan sentimen positif yaitu sebanyak 32.198 review. Sedangkan untuk review dengan sentimen negatif yaitu sebanyak 5.753 review. Selanjutnya disini dibuat diagram dari distribusi sentimen sebagai berikut:



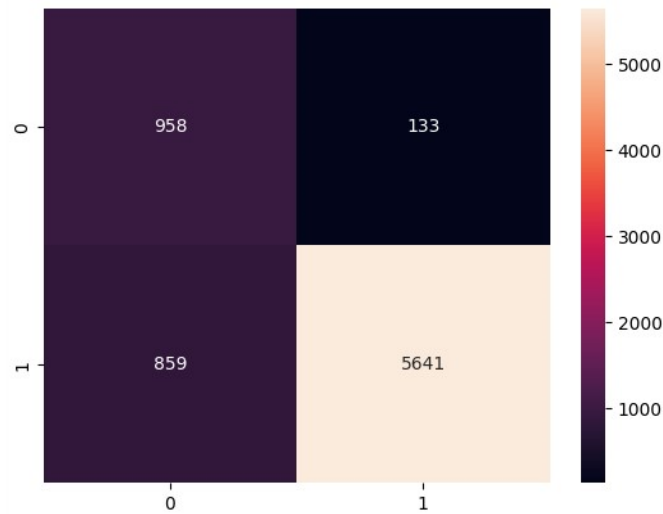
Gambar 4. Diagram Distribusi Sentimen

D. Modeling

Dalam proses modeling produk kami tentunya kita mulai dari membuat EDA hingga Pre-Processing. Selanjutnya proses mengklasifikasikan teks yang dapat dideteksi sentimen nya. Adapun sentiment mapping yang kami buat ada 2 kategori yaitu, Negatif dan Positif. Model Naive Bayes dipilih untuk tugas analisis sentimen teks karena keunggulannya yang sesuai dengan kebutuhan. Model ini mudah diimplementasikan, efisien dalam pemrosesan teks, dan mampu menangani dataset berdimensi tinggi dengan kata-kata sebagai fitur [10]. Selain itu, Naive Bayes memiliki kemampuan untuk menangani data teks dengan lebih dari dua kategori sentimen, serta mampu mengklasifikasikan data kategorikal. Meskipun memiliki asumsi independensi "naif," model ini sering memberikan hasil yang memadai dalam analisis sentimen, menjadikannya pilihan yang populer untuk proyek ini.

E. Evaluation

Metrik evaluasi yang digunakan untuk mengukur kinerja model yaitu akurasi dan F1 Score dikarenakan penggunaan akurasi memberikan gambaran umum tentang kinerja model, sedangkan F1 Score memberikan keseimbangan antara ketepatan dan keberhasilan dalam mengidentifikasi kelas tertentu. Berikut adalah Confusion matrix dan Classification report untuk data latih dan data uji:



Gambar 5. Confusion Matrix Data Uji

Selanjutnya yaitu membuat laporan klasifikasi atau classification report dari data test yang didapatkan hasil sebagai berikut:

	precision	recall	f1-score	support
Kelas 0	0.53	0.88	0.66	1091
Kelas 1	0.98	0.87	0.92	6500
accuracy			0.87	7591
macro avg	0.75	0.87	0.79	7591
weighted avg	0.91	0.87	0.88	7591

Dari *classification report* pada data test di atas dihasilkan bahwa model memiliki kinerja baik pada kelas 1 (positif) dibuktikan dengan nilai precision, recall, dan f1-score yang tinggi. Sedangkan pada kelas 0 (negatif) memiliki nilai recall yang tinggi menunjukkan bahwa model cukup baik dalam mengidentifikasi kelas 0, tetapi nilai precision yang rendah menunjukkan sebagian besar yang diprediksi dari kelas 0 sebenarnya merupakan kelas 1. Dari hasil akurasi menunjukkan bahwa model dapat memprediksi sentimen dari review sebesar 87 %.

Setelah dilakukan classification report pada data test, maka selanjutnya akan dilakukan prediksi dan mengukur akurasi pada data uji dengan hasil Akurasi: 92.20%. Kemudian dilanjutkan membuat confusion matriks pada data uji yaitu sebagai berikut:



Gambar 6. Confusion Matrix Data Latih

Gambar 3. menunjukkan bahwa terdapat 4307 review yang benar sentimen negatif diprediksi dengan benar sebagai sentimen negatif. Terdapat 4290 review yang benar sentimen positif diprediksi benar sebagai positif. Kemudian dilakukan juga Classification Report pada data Uji tersebut yang didapatkan hasil berikut ini:

```
Classification Report - Training Data:
      precision    recall  f1-score   support

0               0.92      0.92      0.92     4662
1               0.92      0.92      0.92     4662

accuracy              0.92     9324
macro avg             0.92      0.92      0.92     9324
weighted avg          0.92      0.92      0.92     9324
```

Model pada data train memiliki kinerja yang sangat baik dengan tingkat akurasi, precision, recall, dan F1-score yang sama tinggi untuk kedua kelas dengan klasifikasi yang seimbang terhadap kelas positif (1) dan negatif (0).

Berdasarkan Classification Report pada data uji dan data latih, maka dihasilkan bahwa model pada data uji memiliki performa yang cukup baik dan tidak terlalu berbeda dibandingkan dengan performa model pada data latih. Hasil ini menunjukkan bahwa model yang dihasilkan cukup baik.

F. Deployment

Aplikasi Smartshop Review berhasil dibuat menggunakan Streamlit di Google Collab. Pengguna dapat memasukkan ulasan produk, dan aplikasi akan memprediksi sentimen ulasan tersebut. Model Naive Bayes yang sudah dilatih dengan data latih digunakan untuk memprediksi sentimen.

G. Tampilan Aplikasi

Aplikasi yang dibuat memiliki tampilan user-friendly dengan form input untuk memasukkan ulasan produk, kemudian menampilkan prediksi sentimen serta contoh ulasan positif dan negatif.



Gambar 7. Tampilan Web Aplikasi SMARTSHOP-REVIEW

Dengan demikian, proyek ini berhasil mengembangkan sebuah aplikasi yang dapat membantu dalam menganalisis sentimen ulasan produk Lazada, memberikan wawasan berharga bagi penjual dan konsumen, serta meningkatkan kemampuan e-commerce dalam memahami kebutuhan pelanggan.

4. KESIMPULAN

- [1] R. L. Atimi and Enda Esyudha Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *J. Sains dan Inform.*, vol. 8, no. 1, pp. 88–96, 2022, doi: 10.34128/jsi.v8i1.419.
- [2] D. Yosmita Praptiwi, "Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine Dan Maximum Entropy," 2018.
- [3] M. E. Alzahrani, T. H. H. Aldhyani, S. N. Alsubari, M. M. Althobaiti, and A. Fahad, "Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3840071.
- [4] K. Norman, Z. Li, Y. T. Oh, G. Golwala, S. Sundaram, and J. Allebach, "Application of natural language processing to an online fashion marketplace," *IS T Int. Symp. Electron. Imaging Sci. Technol.*, pp. 1–5, 2018, doi: 10.2352/ISSN.2470-1173.2018.10.IMAWM-444.
- [5] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, "Corrigendum: Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 2, p. 169, 2018, doi: 10.20473/jisebi.4.2.169.
- [6] A. R. Isnain, N. S. Marga, and D. Alita, "Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 1, p. 55, 2021, doi: 10.22146/ijccs.60718.
- [7] W. Kurnia, "Sentimen Analisis Aplikasi E-Commerce Berdasarkan Ulasan Pengguna Menggunakan Algoritma Stochastic Gradient Descent," *J. Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 138–143, 2023, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/2561%0Ahttp://jim.teknokrat.ac.id/index.php/sisteminformasi/article/download/2561/782>
- [8] A. Ibnu, "Rekayasa Perangkat Lunak Dengan Model Unified Process Studi Kasus: Sistem Informasi Journal," *J. Pilar Nusa Mandiri*, vol. 12, no. 1, p. 11, 2022.
- [9] A. Chamekh, M. Mahfoudh, and G. Forestier, "Sentiment Analysis Based on Deep Learning in E-Commerce," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13369 LNAI, pp. 498–507, 2022, doi: 10.1007/978-3-031-10986-7_40.
- [10] B. Gunawan, H. P. Sasty, and E. P. Esyudha, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *Jepin*, vol. 4, no. 2, pp. 17–29, 2018, [Online]. Available: www.femaledaily.com