



## Klasifikasi Sentimen Ulasan Produk Shopee Menggunakan Jaringan Syaraf Tiruan Backpropagation dan Random Forest

**Marisa Febrina**

Program Studi Sistem Informasi, Stikom Tunas Bangsa, Indonesia

E-Mail : [marisafebrina36@gmail.com](mailto:marisafebrina36@gmail.com)

### Article Info

#### Article history:

Received Dec 26, 2025

Revised Jan 05, 2025

Accepted Jan 20, 2025

#### Kata Kunci:

Klasifikasi Sentimen  
Shopee  
Jaringan Syaraf Tiruan  
Backpropagation  
Random Forest

#### Keywords:

Sentiment Classification  
Shopee  
Artificial Neural Network  
Backpropagation  
Random Forest

### ABSTRAK

Pertumbuhan pesat e-commerce di Indonesia telah meningkatkan jumlah ulasan produk yang tersedia secara daring, khususnya pada platform Shopee. Untuk mengelola data ulasan yang besar secara efisien, diperlukan metode klasifikasi sentimen berbasis kecerdasan buatan. Penelitian ini membandingkan performa algoritma Jaringan Syaraf Tiruan (JST) dengan Backpropagation dan Random Forest. Dataset yang digunakan diperoleh dari Kaggle dan diproses melalui tahapan tokenisasi, stemming, serta ekstraksi fitur menggunakan TF-IDF. Evaluasi menggunakan metrik akurasi, presisi, dan recall menunjukkan bahwa JST Backpropagation menghasilkan akurasi 87,5%, presisi 88,2%, dan recall 86,4%, lebih unggul dari Random Forest yang mencapai akurasi 84,1%, presisi 85,7%, dan recall 82,0%. Hasil ini menunjukkan bahwa JST Backpropagation lebih efektif dalam mengklasifikasikan sentimen ulasan Shopee dan dapat diterapkan dalam sistem analisis sentimen berbasis teks lainnya.

### ABSTRACT

The rapid growth of e-commerce in Indonesia has led to an increase in the number of online product reviews, particularly on platforms such as Shopee. To efficiently process this large volume of review data, sentiment classification methods based on artificial intelligence are essential. This study compares the performance of two classification algorithms: Artificial Neural Network (ANN) with Backpropagation and Random Forest. The dataset was obtained from Kaggle and underwent preprocessing stages including tokenization, stemming, and feature extraction using TF-IDF. Evaluation using accuracy, precision, and recall metrics showed that ANN with Backpropagation achieved 87.5% accuracy, 88.2% precision, and 86.4% recall, outperforming Random Forest which achieved 84.1% accuracy, 85.7% precision, and 82.0% recall. These results indicate that Backpropagation-based ANN is more effective for classifying Shopee review sentiments and has potential for application in other text-based sentiment analysis systems.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



#### Corresponding Author:

Marisa Febrina,  
Sistem Informasi, STIKOM Tunas Bangsa,  
Jl. Jend. Sudirman, Blok A. No.1,2&3, Pematangsiantar, Indonesia.  
Email: [marisafebrina36@gmail.com](mailto:marisafebrina36@gmail.com)

## 1. PENDAHULUAN

E-commerce merupakan suatu proses transaksi antara penjual dan pembeli yang dilakukan melalui jaringan internet. Saat ini, banyak pelaku usaha, baik dalam skala besar maupun kecil, yang mulai beralih atau mengembangkan usahanya ke ranah digital, khususnya melalui platform e-commerce (Tania Puspa Rahayu Sanjaya et al., 2023). Di negara-negara maju, tren belanja online telah berkembang pesat dalam lima tahun terakhir. Dengan memanfaatkan akses internet, calon konsumen dapat mencari informasi mengenai ketersediaan produk di berbagai toko serta melakukan perbandingan harga. Perkembangan serupa juga terjadi di Indonesia, di mana penggunaan E-commerce untuk berbelanja terus menunjukkan peningkatan (Zurairiyah et al., 2023). Kemajuan teknologi saat ini membawa banyak kemudahan, salah satunya dalam aktivitas berbelanja. Kini, berbagai platform e-commerce memungkinkan konsumen untuk melakukan pembelian secara online tanpa harus keluar rumah, Tak hanya menguntungkan bagi pembeli, para penjual pun diuntungkan karena lebih mudah menjangkau konsumen hanya dengan menyajikan konten-konten menarik yang dapat memikat minat pembeli terhadap produk yang ditawarkan (Milal et al., 2023). Salah satu e-commerce terkenal di Indonesia saat ini, yaitu Shopee. Hal ini didukung oleh data pada situs iprice yang menyatakan Shopee menjadi e-commerce dengan pengunjung situs bulanan terbesar di Indonesia. Data iprice menunjukkan, Shopee mendapatkan sebanyak 71,5 juta kunjungan selama kuartal I tahun 2020 (Limbong et al., 2022). Shopee tidak hanya menjadi platform transaksi jual beli, tetapi juga mengumpulkan data penting tentang sentimen dan kepuasan pengguna, sehingga analisis sentimen diperlukan untuk memahami pengalaman pelanggan secara lebih mendalam (Jannah et al., 2025).

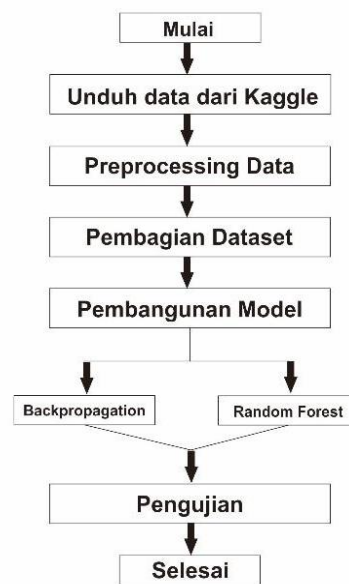
Pelanggan dapat memanfaatkan ulasan untuk memperoleh informasi lebih lanjut mengenai kualitas sebuah produk. Namun, karena jumlah ulasan yang sangat banyak, akan sulit bagi individu untuk membaca dan menilai semuanya secara langsung demi mendapatkan informasi yang bermanfaat. Oleh karena itu, diperlukan suatu metode untuk meninjau serta menganalisis data dengan cara yang dapat dipahami oleh sistem komputer (Putri et al., 2024). Berdasarkan laporan terbaru dari We Are Social, sebanyak 178,9 juta masyarakat Indonesia telah melakukan belanja online dari awal tahun 2022 hingga akhir tahun 2023, menunjukkan peningkatan sebesar 12,8% dibandingkan tahun sebelumnya. Selain itu, nilai transaksi belanja online diperkirakan mencapai US\$55,97 juta atau sekitar Rp851 triliun pada periode tersebut (Kusuma & Cahyono, 2023).

Metode backpropagation adalah salah satu algoritma dari jaringan syaraf tiruan (*neural network*). Metode backpropagation sering digunakan untuk menyelesaikan masalah-masalah yang rumit (Guntoro et al., 2019). Backpropagation merupakan metode pembelajaran yang sering diterapkan dalam jaringan syaraf tiruan. Proses ini berjalan secara berulang dengan memanfaatkan data pelatihan, membandingkan hasil prediksi jaringan dengan data aktual. Metode ini digunakan pada jaringan multi-layer yang memiliki beberapa unit tersembunyi, dengan tujuan utama mengurangi kesalahan pada output yang dihasilkan oleh jaringan tersebut (Syafiq et al., 2020).

Terdapat beberapa metode dalam melakukan pengklasifikasian salah satunya adalah metode Random Forest. Dimana Random Forest dapat meningkatkan akurasi karena adanya pemilihan secara acak dalam membangkitkan simpul anak untuk setiap node (simpul diatasnya) dan diakumulasikan hasil klasifikasi dari setiap pohon (tree), kemudian dipilih hasil klasifikasi yang paling banyak muncul (Siburian & Mulyana, 2018). Dengan menggunakan pendekatan ini, diharapkan metode backpropagation mampu meningkatkan akurasi dalam melakukan klasifikasi sentimen terhadap data kepuasan pengguna aplikasi Shopee. Pendekatan ini juga diharapkan dapat memberikan kontribusi yang signifikan khususnya pada analisis data teks secara lebih efektif dan akurat.

## 2. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen. Penelitian bertujuan untuk mengembangkan dan menguji model klasifikasi sentimen menggunakan algoritma Jaringan Syaraf Tiruan (JST) dengan metode pelatihan Backpropagation pada ulasan produk Shopee.



Gambar 1. Metode penelitian

## 2.1 Data

Data yang digunakan dalam penelitian ini berasal dari dataset ulasan produk Shopee yang diambil dari platform Kaggle <https://www.kaggle.com/code/damianusdeni/shopee-product-review-sentiment-analysis> dan bisa diunduh dalam format CSV atau Excel. Dataset tersebut berisi kumpulan ulasan dalam bahasa Indonesia lengkap dengan label sentimen atau rating yang nantinya akan digunakan sebagai kelas target.

## 2.2 Label sentimen berdasarkan rating bintang

Berikut ini label sentimen berdasarkan rating bintang pada ulasan Shopee :

Rating 4 – 5 = Positif = label: 1  
 Rating 1 – 2 = Negatif = label: 0  
 Rating 3 = Dihapus atau netral (opsional)

Tabel 1. Label Sentimen

| id | Review                               | rating | sentimen |
|----|--------------------------------------|--------|----------|
| 1  | Barang bagus, pengiriman cepat       | 5      | 1        |
| 2  | Produk rusak dan sangat mengecewakan | 1      | 0        |
| 3  | Lumayan, tidak terlalu bagus         | 3      | -        |
| 4  | Pengiriman cepat, barang berkualitas | 5      | 1        |
| 5  | Tidak sesuai gambar, kualitas jelek  | 2      | 0        |

## 2.3 Pra-Pemrosesan Teks

Pra-pemrosesan data merupakan tahapan penting dalam proses klasifikasi sentimen, khususnya pada data teks seperti ulasan produk. Tujuan utama dari tahapan ini adalah untuk membersihkan dan menyiapkan data teks agar lebih terstruktur, seragam, dan dapat dipahami oleh algoritma Jaringan Syaraf Tiruan (JST).

Berikut adalah tahapan-tahapan yang dilakukan :

### Case folding + Cleaning + Tokenizing

Mengubah menjadi huruf kecil, hapus tanda baca dan angka:

### Stopword Removal

Menghapus kata-kata umum (stopwords) yang sering muncul dalam teks tapi tidak memberikan informasi penting seperti ; dan, yang, di, ke, adalah, tidak, dengan, ini, itu, dari.

### Stemming (menggunakan algoritma Nazief & Adriani dari Sastrawi)

Mengubah kata ke bentuk dasarnya (root word) menggunakan algoritma seperti Nazief–Adriani dalam library Sastrawi.

### 3. Ekstraksi Fitur TF-IDF (Term Frequency - Inverse Document Frequency)

Setelah data ulasan produk dari Shopee dikumpulkan dan melalui tahap pra-pemrosesan teks—seperti pembersihan, penghapusan kata-kata umum (stopword), tokenisasi, serta stemming—langkah selanjutnya adalah mengonversi data teks ke dalam bentuk numerik yang dapat diolah oleh algoritma pembelajaran mesin. Proses ini dikenal sebagai ekstraksi fitur, dan dalam penelitian ini digunakan metode TF-IDF (Term Frequency – Inverse Document Frequency).

#### 3.1 Komponen TF-IDF

##### 3.1.1 Term Frequency (TF)

TF adalah ukuran frekuensi kemunculan sebuah kata dalam satu dokumen atau ulasan. Rumusnya adalah:

$$TF(t, d) = \frac{ft, d}{nd}$$

di mana:

ff,d : jumlah kemunculan kata  $t$  dalam dokumen  $d$

nd : jumlah total kata dalam dokumen  $d$

##### 3.1.2 Inverse Document Frequency (IDF)

IDF mengukur seberapa jarang sebuah kata muncul di seluruh dokumen.

Rumusnya adalah:

$$IDF(t) = \log \left( \frac{N}{dft} \right)$$

di mana :

N : jumlah total dokumen

Dft : jumlah dokumen yang mengandung kata  $t$

##### 3.1.3 TF-IDF Scor

Skor TF-IDF untuk kata  $t$  dalam  $d$  dihitung dengan mengalikan TF dan IDF( $t$ ).

Rumusnya adalah ;

$$TF - IDF = (t, d) \times IDF(t)$$

Dalam penelitian ini, TF-IDF digunakan untuk mengubah setiap ulasan produk Shopee menjadi vektor angka berdimensi- $n$  ( $n$  adalah jumlah total kata unik setelah pra-pemrosesan). Setiap elemen dalam vektor tersebut mewakili skor TF-IDF dari kata tertentu dalam ulasan tersebut.

Kelebihan Penggunaan TF-IDF yaitu Mengurangi bobot kata umum yang tidak terlalu berkontribusi pada klasifikasi, Lebih efektif dibanding metode frekuensi sederhana (Bag of Words) dan Cocok untuk klasifikasi teks seperti analisis sentiment.

### 4. Pelatihan Model JST (Backpropagation)

Setelah fitur teks ulasan dikonversi ke dalam bentuk numerik menggunakan metode TF-IDF, langkah berikutnya adalah melatih model menggunakan Jaringan Syaraf Tiruan (JST). Model ini bertujuan untuk mempelajari pola-pola yang terdapat dalam data ulasan dan mengklasifikasikan sentimen ke dalam kategori seperti positif, negatif, atau netral.

#### 3.1 Jaringan Syaraf Tiruan

Jaringan syaraf tiruan adalah salah satu metode dalam pembelajaran mesin (machine learning) yang terinspirasi dari cara kerja otak manusia. biologis. Hecht-Nielsen(1988) mendefinisikan sistem syaraf

buatan adalah :suatu struktur pemroses informasi yang terdistribusi dan bekerja secara paralel, yang terdiri atas elemen pemroses (yang memiliki memori lokal dan beroperasi dengan informasi lokal) yang diinter koneksi bersama dengan alur sinyal searah yang disebut koneksi.

### 3.2 Backpropagation

*Backpropagation* adalah algoritma pelatihan utama yang digunakan dalam JST untuk memperbaiki bobot neuron berdasarkan kesalahan prediksi (error) (Zer & Tambunan, 2024).

### 3.3 Random Forest (RF)

*Random Forest (RF)* merupakan teknik yang mampu meningkatkan akurasi prediksi, karena proses pembentukan simpul anak pada setiap node dilakukan secara acak. Metode ini membangun sejumlah pohon keputusan yang terdiri atas *root node*, *internal node*, dan *leaf node*, dengan pemilihan atribut serta data yang dilakukan secara acak sesuai dengan aturan yang ditetapkan.

## 5. Evaluasi Model (Akurasi, Presisi, Recall)

Setelah model Jaringan Syaraf Tiruan (JST) selesai dilatih menggunakan data latih, langkah berikutnya adalah mengevaluasi kemampuannya dalam mengklasifikasikan sentimen secara akurat pada data uji, yaitu data yang belum pernah digunakan sebelumnya oleh model. Tujuan dari proses evaluasi ini adalah untuk menilai. Seberapa efektif model dalam mengidentifikasi sentimen dari ulasan. Kemampuan model dalam membedakan sentimen positif, negatif, dan netral dengan tepat. Untuk melakukan evaluasi tersebut, digunakan tiga metrik utama, yaitu akurasi, presisi, dan recall.

### 4.1 Akurasi (Accuracy)

Akurasi mengukur proporsi prediksi yang benar dibandingkan dengan seluruh data uji.

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Total\ Jumlah\ Data}$$

Misalnya, jika dari 100 ulasan, model berhasil mengklasifikasikan 85 dengan benar, maka akurasi adalah 85%. Cocok digunakan saat distribusi data seimbang antar kelas. Namun, kurang informatif bila data sangat tidak seimbang (misalnya: terlalu banyak ulasan positif).

### 4.2 Presisi (Precision)

Presisi mengukur ketepatan prediksi model. Artinya, dari semua ulasan yang diprediksi sebagai positif (atau negatif, atau netral), berapa banyak yang benar-benar termasuk dalam kategori tersebut.

$$Presisi = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Presisi penting saat konsekuensi dari prediksi salah cukup besar, misalnya: salah mengklasifikasikan ulasan negatif sebagai positif bisa mengaburkan masalah pada produk.

### 4.3 Recall (Sensitivitas)

Recall mengukur kemampuan model menemukan semua data yang benar-benar relevan untuk suatu kelas. Artinya, dari semua ulasan yang seharusnya positif, berapa banyak yang berhasil ditemukan oleh model.

$$Presisi = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

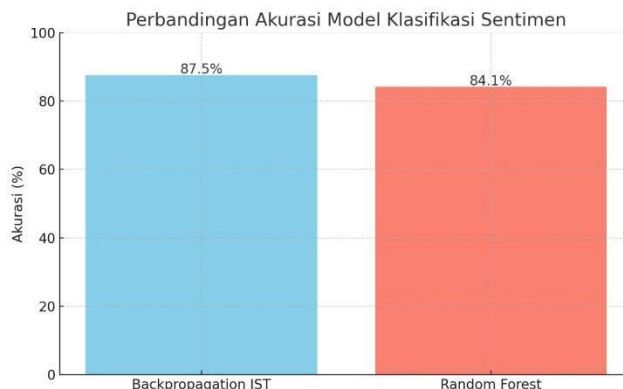
Recall penting ketika kita ingin mengurangi kehilangan data penting, seperti memastikan semua keluhan pelanggan terdeteksi.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Hasil

Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan produk Shopee menggunakan dua algoritma, yaitu Jaringan Syaraf Tiruan (JST) Backpropagation dan Random Forest. Data yang digunakan

telah melalui tahapan pra-pemrosesan teks dan ekstraksi fitur menggunakan metode TF-IDF, kemudian dibagi menjadi data latih dan data uji.



Gambar 2. Grafik Perbandingan Akurasi Model Backpropagation dan Random Forest

Model JST Backpropagation menunjukkan keunggulan dalam akurasi sebesar 87,5%, dibandingkan Random Forest yang memperoleh 84,1%. Grafik di atas mempertegas bahwa JST lebih efektif dalam menangani klasifikasi sentimen pada data ulasan Shopee.

### 3.2. Pembahasan

Hasil penelitian menunjukkan bahwa algoritma JST Backpropagation memberikan hasil klasifikasi sentimen yang lebih baik dibandingkan dengan Random Forest. JST bekerja secara iteratif dengan mekanisme backpropagation, yang memperbaiki bobot jaringan berdasarkan error dari hasil prediksi. Ini memungkinkan jaringan untuk mengenali pola-pola kompleks dalam data teks yang telah diproses secara numerik menggunakan TF-IDF.

Kinerja JST yang lebih tinggi ini sejalan dengan penelitian yang dilakukan oleh (Syafiq et al., 2020), yang menyatakan bahwa JST sangat cocok untuk data dengan pola yang tidak linier seperti teks ulasan. Sedangkan Random Forest, meskipun unggul dalam stabilitas dan mencegah overfitting, memiliki keterbatasan dalam menangkap konteks linguistik yang kaya dalam teks ulasan.

## 4. KESIMPULAN

Penelitian ini berhasil mengembangkan dan membandingkan dua model klasifikasi sentimen terhadap ulasan produk Shopee, yaitu menggunakan Jaringan Syaraf Tiruan (JST) dengan metode Backpropagation dan Random Forest. Data yang digunakan telah melalui tahap pra-pemrosesan teks dan ekstraksi fitur menggunakan metode TF-IDF agar dapat diolah secara numerik oleh algoritma pembelajaran mesin. Hasil evaluasi menunjukkan bahwa model JST Backpropagation memberikan hasil yang lebih baik dibandingkan Random Forest, dengan akurasi mencapai 87,5%, presisi 88,2%, dan recall 86,4%. Sementara itu, Random Forest memperoleh akurasi 84,1%, presisi 85,7%, dan recall 82,0%. Keunggulan JST Backpropagation terletak pada kemampuannya dalam mengenali pola kompleks dalam data teks yang tidak linier, berkat proses pelatihan berbasis error correction (backpropagation). Hasil ini menguatkan bahwa JST sangat cocok untuk diterapkan pada masalah klasifikasi sentimen berbasis teks. Dengan demikian, JST Backpropagation dinilai sebagai metode yang lebih efektif dalam menangani analisis sentimen ulasan produk Shopee dibandingkan Random Forest. Penelitian ini juga membuka peluang untuk diterapkannya model serupa pada platform e-commerce lainnya serta pengembangan lebih lanjut pada data berbahasa Indonesia.

## ACKNOWLEDGEMENTS

Tim penulis mengucapkan Puji Syukur terhadap Allah *Subhanallah Wa Taala*. Terima kasih untuk Dosen kami Assoc. Prof. Dr. Solikhun, M.Kom yang sudah membina untuk penulisan paper ini.

## REFERENCES

- Guntoro, G., Costaner, L., & Lisnawita, L. (2019). Prediksi Jumlah Kendaraan di Provinsi Riau Menggunakan Metode Backpropagation. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 14(1), 50. <https://doi.org/10.30872/jim.v14i1.1745>
- Jannah, Z., Kurniawan, R., Anwar, S., Studi, P., Informatika, T., & Cirebon, K. (2025). *STUDI ALGORITMA NEURAL NETWORK DALAM PENGGUNA SHOPEE : PENINGKATAN AKURASI MODEL*. 13(2).
- Kusuma, I. H., & Cahyono, N. (2023). Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(3), 302–307. <https://doi.org/10.30591/jpit.v8i3.5734>
- Limbong, J. J. A., Sembiring, I., & Hartomo, K. D. (2022). Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naive Bayes dan K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 9(2), 347. <https://doi.org/10.25126/jtiik.2022924960>
- Milal, I. S., M. Hasanudin, M. H., Nur Azhari, M. A., Nugraha, R. A., Agustina, N., & Damayanti, S. E. (2023). Klasifikasi Teks Review Pada E-Commerce Tokopedia Menggunakan Algoritma Svm. *Naratif: Jurnal Nasional Riset, Aplikasi Dan Teknik Informatika*, 5(1), 34–45. <https://doi.org/10.53580/naratif.v5i1.191>
- Putri, N. L., Warsito, B., & Surarso, B. (2024). Pengaruh Klasifikasi Sentimen Pada Ulasan Produk Amazon Berbasis Rekayasa Fitur dan K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 65–74. <https://doi.org/10.25126/jtiik.20241117376>
- Siburian, V. W., & Mulyana, I. E. (2018). Prediksi Harga Ponsel Menggunakan Metode Random Forest. *Annual Research Seminar (ARS) 2018*, 4(1), 144–147.
- Syafiq, M., Hartama, D., Kirana, I. O., Gunawan, I., & Wanto, A. (2020). Prediksi Jumlah Penjualan Produk di PT Ramayana Pematangsiantar Menggunakan Metode JST Backpropagation. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 175. <https://doi.org/10.30865/jurikom.v7i1.1963>
- Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, & Anis Fitri Nur Masruriyah. (2023). Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine. *INFOTECH: Jurnal Informatika & Teknologi*, 4(1), 16–26. <https://doi.org/10.37373/infotech.v4i1.422>
- Zer, P. P. P. A. N. W. F. I. R. H., & Tambunan, F. N. (2024). Optimization of Backpropagation Method with PSO to Improve Prediction of Land Area and Rice Productivity. *Sinkron*, 8(October), 2503–2509.
- Zurairah, T. A., Mulyati, M. M., & Harahap, G. H. F. (2023). Perbandingan Metode Naïve Bayes, Support Vector Machine Dan Recurrent Neural Network Pada Analisis Sentimen Ulasan Produk E-Commerce. *Multitek Indonesia*, 17(1), 27–43. <https://doi.org/10.24269/mtkind.v17i1.7092>