

ANALISIS SENTIMEN MASYARAKAT TERHADAP PINJAMAN ONLINE DI APLIKASI X MENGGUNAKAN LONG SHORT-TERM MEMORY

*Analysis of Public Sentiment Towards Online Loans in Application X
Using LongShort-Term Memory*

Hafizh Maalik Falah^{1*}, Castaka Agus Sugianto²

¹Program Studi Teknik Informatika, ²Politeknik TEDC Bandung

*Penulis Korespondensi: hafizhmf23@gmail.com

DOI: <https://doi.org/10.58217/ipsikom.v13i2.95>

ABSTRAK

The development of online loans in Indonesia has led to various public opinions spread across social media, one of which is the X platform. This research aims to analyze public sentiment towards online loans using the Long Short-Term Memory (LSTM) method. The data used consists of 702 Indonesian tweets collected through a crawling process with Tweet Harvest. Of these, 480 tweets were classified as positive sentiment and 222 as negative. The research process includes preprocessing, manual labeling, model training, and evaluation stages. The model was built using Sequential architecture from Keras, consisting of embedding layer, LSTM layer 128 units, 30% dropout, and output layer with softmax activation function. The model was trained using 562 tweets as training data and 140 tweets as validation data with a ratio of 80:20, for 10 epochs and batch size 64. The final evaluation using the entire dataset resulted in 92.59% accuracy, with 79.06% precision, 79.43% recall, and 79.14% F1-score. These results show that LSTM is able to classify sentiment stably and effectively, and has strong potential in sentiment analysis on short text data such as tweets.

Keywords: Analisis, LSTM, Pinjaman, Sentimen,

PENDAHULUAN

Kemajuan teknologi di era digital, khususnya dalam revolusi industri 4.0, mendorong hadirnya *Financial Technology (Fintech)* yang menawarkan layanan keuangan efisien dan aman. Salah satu layanan *Fintech* yang berkembang pesat adalah pinjaman *online* (pinjol), yang memberikan akses kredit secara cepat dan praktis (Alif, Alamsyah, and Arif 2024). Pinjaman *online* merupakan inovasi digital dalam layanan peminjaman uang berbasis daring yang menawarkan kemudahan akses kredit secara cepat melalui aplikasi dan *platform* dengan ketentuan pembayaran yang disepakati secara *online* (Afandi 2020). Pinjaman *online* berkembang pesat di Indonesia karena proses administrasinya lebih sederhana dibandingkan pinjaman bank (Lestari 2022). Masyarakat kerap tergoda menggunakan pinjaman *online*, namun tanpa disadari hal ini memicu masalah finansial akibat kewajiban membayar kembali dengan bunga tinggi (Kartikawati and Soediro 2022). Situasi ini memunculkan berbagai opini masyarakat yang banyak diungkapkan melalui media sosial X. X merupakan salah satu media

sosial terpopuler di internet yang berperan sebagai sarana komunikasi di masyarakat (Sugianto and Tresnawati 2022). Pengguna X dapat membagikan pandangan atau opini melalui *tweet* berdurasi singkat hingga 140 karakter, menjadikannya platform yang menyajikan beragam opini publik dari seluruh dunia (Widowati and Sadikin 2021). Untuk memahami opini masyarakat terhadap pinjaman online di X, digunakan analisis sentimen berdasarkan data teks seperti *tweet*.

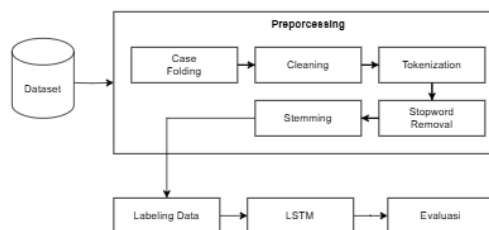
Analisis sentimen merupakan bagian dari *Natural Language Processing (NLP)* yang bertujuan mengidentifikasi dan mengklasifikasikan opini atau emosi dalam teks menjadi sentimen positif, negatif, atau netral (Riansyah and Sugianto 2024). Penelitian ini menerapkan analisis sentimen untuk mengidentifikasi respons masyarakat terhadap pinjaman online menggunakan metode *Deep Learning*, yaitu *Long Short-Term Memory (LSTM)* (Rahman, Sari, and Yudistira 2021). LSTM efektif untuk klasifikasi teks karena mampu menangkap hubungan jangka panjang dalam data berurutan, dan telah

terbukti menghasilkan akurasi tinggi dalam klasifikasi sentimen *tweet* (Shah and Li 2021).

Sejumlah penelitian terdahulu juga menunjukkan keunggulan LSTM dalam analisis sentimen. Pipin dan Kurniawan, misalnya, berhasil mengklasifikasikan 658 *tweet* terkait kebijakan Merdeka Belajar Kampus Merdeka (MBKM) menggunakan LSTM dengan akurasi mencapai 80,42% (Pipin and Kurniawan 2022). Penelitian lain oleh Rahman dan rekan-rekannya menunjukkan bahwa LSTM memiliki kinerja lebih baik dibandingkan *Naïve Bayes* dalam menganalisis sentimen ulasan novel, dengan akurasi, presisi, dan *recall* sebesar 83,33%, lebih unggul 1,33% dari *Naïve Bayes* yang hanya mencapai 82% (Rahman et al. 2021). Hasil serupa juga diperoleh dalam studi oleh Isnain rekan-rekannya yang menunjukkan keunggulan LSTM dengan akurasi yang sama, yakni 83,33%, dibandingkan *Naïve Bayes* yang mencapai 82% (Isnain et al. 2022).

METODOLOGI PENELITIAN

Penulis menggunakan metodologi penelitian yang terdiri dari 5 tahapan, yaitu crawling datasaset, preprocessing, labeling data, pelatihan model menggunakan LSTM, dan evaluasi kinerja model. Alur proses penelitian ini divisualisasikan pada Gambar 1.



Gambar 1. Alur penelitian

Dataset

Dataset yang digunakan dalam penelitian ini adalah kumpulan *tweet* yang diperoleh melalui *Tweet Harvest*, dengan jumlah 702 *tweet* yang telah dikumpulkan dan diproses dengan kata kunci “pinjaman *online*”. Pengambilan data dilakukan dalam rentang waktu 1 Januari 2024 hingga 31 Desember 2024.

Preprocessing

Tahapan *preprocessing* dalam penelitian ini meliputi lima langkah utama. Pertama, *case folding* dilakukan dengan mengubah seluruh huruf menjadi kecil. Kedua, *cleaning* menghapus karakter khusus, angka, dan simbol yang tidak relevan. Ketiga, *tokenized* memecah teks menjadi kata-kata terpisah. Keempat, *stopword removal* menghapus kata umum seperti “yang” dan “itu” menggunakan *library* Sastrawi. Terakhir, *stemming* mengubah kata ke bentuk dasar dengan menghilangkan imbuhan.

Labeling Data

Setelah *preprocessing*, *tweet* diberi label secara manual menggunakan *Microsoft Excel* dengan menilai konteks isi setiap *tweet*. Pelabelan dilakukan ke dalam dua kategori, yaitu positif dan negatif, untuk memastikan akurasi dan relevansi data sebelum digunakan dalam pelatihan model.

Long Short-Term Memory (LSTM)

Setelah proses pelabelan selesai, tahap selanjutnya adalah melakukan klasifikasi sentimen menggunakan LSTM. LSTM dipilih karena kemampuannya dalam mengolah data berurutan dan mempertahankan informasi penting dalam konteks teks, sehingga cocok digunakan untuk menganalisis opini dalam *tweet*. Model dibangun menggunakan arsitektur *Sequential* dari Keras yang terdiri dari *embedding layer*, *LSTM layer*, *dropout*, dan *output layer* dengan aktivasi *softmax* untuk klasifikasi dua kelas, yaitu positif dan negatif.

Evaluasi

Evaluasi dilakukan untuk mengukur performa model LSTM dalam mengklasifikasikan sentimen positif dan negatif serta memastikan kemampuannya memprediksi data yang belum pernah dilihat sebelumnya. *Accuracy*, yang merupakan proporsi prediksi benar terhadap total data uji, digunakan sebagai indikator awal performa model. Selanjutnya, evaluasi diperkuat dengan *confusion matrix* yang mencakup *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang menjadi dasar perhitungan presisi, *recall*, dan *F1-score* untuk menilai ketepatan klasifikasi secara lebih menyeluruh.

$$Accuracy = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Data Uji}} \times 100\%$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{FP + FN}$$

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

HASIL DAN PEMBAHASAN

Dataset

Data yang digunakan terdiri dari 702 sampel berbahasa Indonesia yang diperoleh melalui proses *crawling* otomatis dari platform X menggunakan *tweet harvest* sebagai alat pengambilan data menggunakan kata kunci "pinjaman online". Visualisasi proses *crawling* ditunjukkan pada Gambar 2.

```
# Crawl Data
filename = 'pinjaman_online.csv'
search_keyword = 'pinjaman online since:2024-01-01 until:2024-12-31 lang:id'
limit = 700

!pip install tweet-harvest@2.6.1 -o "[filename]" -s "[search_keyword]" --tab "LATEST" -l [limit] --token [twitter_auth_token]
!python3 tweet-harvest [v2.6.1]
```

Gambar 2. Proses *crawling* data

Preprocessing

Tahap *preprocessing* merupakan proses penting dalam pengolahan data yang bertujuan untuk membersihkan, menyiapkan, dan mengatur data mentah yang diperoleh dari X umumnya masih belum teratur, agar siap digunakan dalam analisis lebih lanjut. Pada tahap ini, data dari file *csv* dibaca dan diperiksa strukturnya untuk memastikan bahwa semua kolom dan isinya terbaca dengan benar. Proses ini memastikan bahwa data yang akan digunakan memiliki kualitas dan struktur yang baik, sehingga mendukung hasil analisis yang akurat dan efisien. *Preprocessing* mencakup beberapa tahapan, seperti *case folding*, *cleaning*, *tokenized*, *stopword removal*, dan *stemming*. Hasil dari tahapan ini dapat dilihat pada (Tabel 1-5).

Tabel 1. Hasil *case folding*

Sebelum	Sesudah
---------	---------

Teruntuk sahabat online ku siapa tau ada yang bisa menolong. Posisi saya di bekasi. Sedang butuh uang pinjaman saya sedang kena musibah yg tak terduga kalau ada temen yg mau membantu saya mau gadai bpkb saja untuk jaminan karena butuhnya cepat malam ini. Nanti ditambah bunga gpp	teruntuk sahabat online ku siapa tau ada yang bisa menolong. posisi saya di bekasi. sedang butuh uang pinjaman saya sedang kena musibah yg tak terduga kalau ada temen yg mau membantu saya mau gadai bpkb saja untuk jaminan karena butuhnya cepat malam ini. nanti ditambah bunga gpp
---	---

Tabel 1 menunjukkan bahwa proses *case folding* berhasil mengubah seluruh teks menjadi huruf kecil, sehingga penyajian data menjadi lebih konsisten dan siap untuk tahap analisis berikutnya.

Tabel 2. Hasil *cleaning*

Sebelum	Sesudah
begini lho hak dan kewajiban nasabah fintech terutama di industri pinjaman online pahami sebelum menggunakan ya ! https://t.co/1AQdHJorL	begini lho hak dan kewajiban nasabah fintech terutama di industri pinjaman online pahami sebelum menggunakan ya

Tabel 2 menunjukkan bahwa proses *cleaning* berhasil menghapus tanda baca tidak perlu dan huruf yang berulang, sehingga teks menjadi lebih sederhana untuk analisis selanjutnya.

Tabel 3. Hasil *tokenized*

Sebelum	Sesudah
bagus banget nih web series bisa menghindari kesadaran tentang pinjaman online	bagus, banget, nih, web, series, bisa, menghindari, kesadaran, tentang, pinjaman, online

Tabel 3 memperlihatkan bahwa proses *tokenized* berhasil memecah kalimat menjadi kata-kata terpisah dalam bentuk *list*, sehingga mempermudah tahapan analisis berikutnya.

Tabel 4. Hasil *stopword removal*

Sebelum	Sesudah
as a person yang menolak paylaterpinjaman online maaf ya wonwoo kita ketemu di lain waktu	as, a, person, menolak, paylaterpinjaman ,online ,maaf, wonwoo, ketemu, waktu

Tabel 4 menunjukkan bahwa proses *stopword removal* berhasil menghapus kata-kata umum dan menyisakan kata-kata yang relevan.

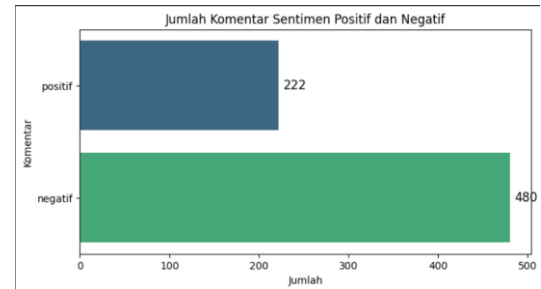
Tabel 5. Hasil *stemming*

Sebelum	Sesudah
suruh pinjol buat lunasin utang dia ke lu lu aja ancur nama di pinjaman online walaupun ojk kalo gak mau ancem kasusin ke keluarga kek kerjaan kek kampus kek	suruh, pinjol, lunasin, utang, lu, lu, aja, ancur, nama, pinjam, online, ojk, kalo, gak, ancem, kasusin, keluarga, kek, kerja, kek, kampus, kek

Tabel 5 menunjukkan bahwa proses *stemming* berhasil menyederhanakan kata ke bentuk dasarnya, seperti "pinjaman" menjadi "pinjam" dan "kerjaan" menjadi "kerja", guna mengurangi variasi kata dalam analisis.

Labeling Data

Setelah melalui tahap preprocessing untuk membersihkan data teks, proses pelabelan dilakukan secara manual. Data yang terdiri dari 702 *tweet* dalam bahasa Indonesia yang berisi pendapat atau reaksi pengguna terhadap pinjaman *online*. Untuk tujuan pelatihan model awal, *tweet-tweet* tersebut diberi *labeling* dalam dua kategori sentimen positif dan negatif. Jumlah komentar dari kategori sebanyak 480 dan negatif 222 dapat dilihat pada Gambar 3.

**Gambar 3.** Jumlah komentar sentimen positif dan negatif

Long Short-Term Memory (LSTM)

Setelah melalui tahap preprocessing dan pelabelan, data digunakan untuk membangun model analisis sentimen menggunakan *Long Short-Term Memory* (LSTM). Model dirancang dengan arsitektur *Sequential* dari *TensorFlow Keras*, terdiri dari *embedding layer* untuk merepresentasikan kata, *LSTM layer* dengan 128 unit untuk menangkap pola urutan, *dropout* 30% untuk mencegah *overfitting*, dan *dense layer* dengan aktivasi *softmax* untuk mengklasifikasikan sentimen menjadi dua kelas: positif dan negatif. LSTM dipilih karena kemampuannya dalam mengolah data berurutan dan mempertahankan informasi penting dalam teks pendek seperti *tweet*, sebagaimana dapat dilihat pada Gambar 4.

Model: "sequential"		
Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 160, 128)	118, 784
lstm (LSTM)	(None, 128)	131, 488
dropout (Dropout)	(None, 128)	0
dense (Dense)	(None, 2)	258
Total params: 442, 620 (1.69 MB)		
Trainable params: 442, 620 (1.69 MB)		
Non-trainable params: 0 (0.00 B)		

Gambar 4. Arsitektur model keras *sequential*

Setelah model selesai dibangun, pelatihan dilakukan menggunakan 562 *tweet* sebagai data latih dan 140 *tweet* sebagai data validasi dengan rasio 80:20. Model dilatih selama 10 *epoch* dengan *batch size* 64, dan performanya tercatat dalam objek history. Hasil pelatihan menunjukkan akurasi mencapai 95% tanpa indikasi *overfitting* yang signifikan, sebagaimana ditampilkan pada Gambar 5.

```

Epoch 1/10
2s/1m/step accuracy: 0.5576 loss: 0.5375 val_accuracy: 0.6089 val_loss: 0.5060
Epoch 2/10
2s/1m/step accuracy: 0.7278 loss: 0.3467 val_accuracy: 0.7873 val_loss: 0.4812
Epoch 3/10
2s/1m/step accuracy: 0.8088 loss: 0.3485 val_accuracy: 0.7738 val_loss: 0.4758
Epoch 4/10
2s/1m/step accuracy: 0.8956 loss: 0.2611 val_accuracy: 0.7943 val_loss: 0.4625
Epoch 5/10
2s/1m/step accuracy: 0.9267 loss: 0.2081 val_accuracy: 0.8227 val_loss: 0.4923
Epoch 6/10
2s/1m/step accuracy: 0.9244 loss: 0.1714 val_accuracy: 0.8227 val_loss: 0.5352
Epoch 7/10
2s/1m/step accuracy: 0.9368 loss: 0.1656 val_accuracy: 0.8085 val_loss: 0.6283
Epoch 8/10
2s/1m/step accuracy: 0.9506 loss: 0.1385 val_accuracy: 0.7872 val_loss: 0.6819
Epoch 9/10
2s/1m/step accuracy: 0.9444 loss: 0.1525 val_accuracy: 0.7738 val_loss: 0.6884
Epoch 10/10
2s/1m/step accuracy: 0.9514 loss: 0.1214 val_accuracy: 0.8018 val_loss: 0.8155

```

Gambar 5. Pelatihan 10 epoch

Untuk mengevaluasi kinerja keseluruhan model, model tersebut juga diuji ulang menggunakan seluruh *dataset* (data pelatihan dan data uji yang digabungkan). Hasil evaluasi menggunakan `model.evaluate(X, y)` menunjukkan akurasi keseluruhan sebesar 92,59%, yang menunjukkan bahwa model tersebut dapat generalisasi dengan baik pada data yang tersedia dapat dilihat pada Gambar 6.

```

# Menggabungkan seluruh data (training + test)
loss, accuracy = model.evaluate(X, y, verbose=0)

print(f"Akurasi Keseluruhan Model: {accuracy * 100:.2f}%")

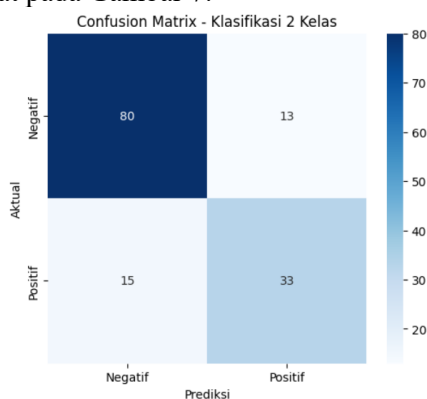
```

Akurasi Keseluruhan Model: 92.59%

Gambar 6. Akurasi keseluruhan model

Evaluasi

Evaluasi model diperkuat dengan visualisasi *confusion matrix* yang menunjukkan performa klasifikasi tiap kelas. Dari 93 *tweet* negatif, 80 diklasifikasikan dengan benar dan 13 salah sebagai positif. Sementara itu, dari 48 *tweet* positif, hanya 33 diklasifikasikan dengan tepat, dan 15 salah sebagai negatif. Hasil ini mengindikasikan bahwa model lebih akurat dalam mengenali sentimen negatif, namun kurang konsisten pada sentimen positif, kemungkinan akibat ketidakseimbangan data dan kompleksitas makna dalam opini positif. Hasil ini dapat dilihat pada Gambar 7.



Gambar 7. Confusion matrix

Model menunjukkan kinerja yang cukup baik dalam klasifikasi dua kelas, yaitu sentimen positif dan negatif. Berdasarkan hasil evaluasi, model memperoleh nilai presisi sebesar 79,06%, *recall* 79,43%, dan *F1-score* 79,14%. Nilai ini mencerminkan bahwa model mampu mengklasifikasikan sentimen secara seimbang dan konsisten. Secara keseluruhan, model memiliki performa yang stabil dalam mengenali pola sentimen, sebagaimana ditampilkan pada Gambar 8.

```

Precision: 0.7905815468196186
Recall: 0.7943262411347518
F1 Score: 0.7913679280505614

```

Gambar 8. Hasil evaluasi

Ketidakseimbangan jumlah data antar kelas, dengan dominasi *tweet* sentimen negatif, serta ukuran dataset yang relatif kecil (702 *tweet*), menjadi faktor utama yang membatasi kemampuan model dalam mengenali variasi sentimen, khususnya untuk kelas positif. Kondisi ini menyebabkan model lebih mudah mengidentifikasi pola dari kelas mayoritas, namun kurang optimal pada kelas minoritas. Meskipun demikian, model tetap mampu mengenali pola sentimen secara umum dengan akurasi di atas 79%, menunjukkan bahwa arsitektur LSTM memiliki potensi yang kuat dalam klasifikasi sentimen pada data teks pendek seperti *tweet*.

KESIMPULAN

Analisis sentimen terhadap 702 *tweet* terkait pinjaman online yang diperoleh melalui proses *crawling* di *Google Colab*, ditemukan bahwa jumlah *tweet* dengan sentimen positif sebanyak 480 data, sedangkan sentimen negatif sebanyak 222 data. Hasil ini menunjukkan bahwa opini masyarakat terhadap pinjaman online di platform X didominasi oleh sentimen positif. *Long Short-Term Memory* (LSTM) yang digunakan dalam klasifikasi sentimen menunjukkan performa yang cukup baik, dengan akurasi 92,59%. Nilai presisi tercatat sebesar 79,06%, *recall* 79,43%, dan *F1-score* 79,14%, dengan kinerja yang lebih unggul dalam mengenali sentimen negatif. Saran untuk penelitian selanjutnya dapat melakukan membandingkan model dengan algoritma klasifikasi lain seperti *Support Vector Machines* (SVM), *Random Forest*, atau *BERT*.

guna memperoleh hasil evaluasi yang lebih menyeluruh dan maksimal.

DAFTAR PUSTAKA

- Afandi, Rival. 2020. "Analisis Sentimen Masyarakat Terhadap Pinjaman Online Di Twitter Menggunakan Algoritma Nbc Dan K-Nn." *Jurnal Ekonomi Volume 18, Nomor 1 Maret 201* 2(1):1–10.
- Alif, Muhammad, Muslim Alamsyah, and Mochammad Firman Arif. 2024. "Analisis Sentimen Twitter Tentang Pinjaman Online Di Indonesia Menggunakan Metode Random Forest." *Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi* 1410–21.
- Isnain, Auliya Rahman, Heni Sulistiani, Bagus Miftaq Hurohman, Andi Nurkholis, and Styawati Styawati. 2022. "Analisis Perbandingan Algoritma LSTM Dan Naive Bayes Untuk Analisis Sentimen." *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)* 8(2):299–303.
- Kartikawati, Ratna, and Soediro Soediro. 2022. "Kontroversi Pinjaman Online Ditinjau Dari Perspektif Moral." *Kosmik Hukum* 22(3):258–67.
- Lestari, Tri Putri. 2022. "Analisis Text Mining Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine (SVM) Dan Social Network Analysis (SNA)." *Jurnal Informatika Ekonomi Bisnis* 4(3):65–71.
- Pipin, Sio, and Heru Kurniawan. 2022. "Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat Di Twitter Menggunakan LSTM." *Jurnal SIFO Mikroskil* 23(2):197–208.
- Rahman, Muhammad Zaini, Yuita Arum Sari, and Novanto Yudistira. 2021. "Analisis Sentimen Tweet COVID-19 Menggunakan Word Embedding Dan Metode Long Short-Term Memory (LSTM)." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 5(11):5120–27.
- Riansyah, Gery, and Castaka Agus Sugianto. 2024. "Analisis Sentimen Ulasan Aplikasi Dana Di Google Play Store Menggunakan Algoritma Naïve Bayes." *JATI (Jurnal Mahasiswa Teknik Informatika)* 8(5):9849–57.
- Shah, Dhruval, and Yanyan Li. 2021. "Twitter Based Sentiment Analysis of Each Presidential Candidate Using Long Short-Term Memory." *Ahmad Hadaegh International Journal of Computer Science and Security (IJCSS)* (15):87–96.
- Sugianto, Castaka Agus, and Shandy Tresnawati. 2022. "A Covid-19 Sentiment Analysis on Twitter Using K-Nearest." 7(1):58–69.
- Widowati, Tanthy Tawaqalia, and Mujiono Sadikin. 2021. "Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naive Bayes Dan Support Vector Machine." *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer* 11(2):626–36.