

## ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP TECH WINTER PADA TWITTER MENGUNAKAN NATURAL LANGUAGE PROCESSING

Rifky Aziz Fauzianto<sup>1</sup>, Supatman<sup>2</sup>

<sup>1,2</sup>Program Studi Informatika, Universitas Mercubuana Yogyakarta

Email: <sup>1</sup>[17111042@students.mercubuana-yogya.ac.id](mailto:17111042@students.mercubuana-yogya.ac.id), <sup>2</sup>[supatman@mercubuana-yogya.ac.id](mailto:supatman@mercubuana-yogya.ac.id)

### Abstrak:

Sektor teknologi secara berkala mengalami fluktuasi, dan istilah "Tech Winter" telah diciptakan untuk menggambarkan periode penurunan. Studi ini bertujuan untuk menganalisis sentimen masyarakat terhadap Tech Winter di Twitter, memanfaatkan kemampuan Natural Language Processing (NLP). Dengan menggunakan berbagai model NLP—Regresi Logistik, Support Vector Machine (SVM), Random Forest, Neural Network, dan Naive Bayes—penelitian ini membandingkan efektivitas mereka dalam menafsirkan sentimen yang diungkapkan dalam tweet. Temuan menawarkan pemahaman tentang opini publik selama periode ini dan mengungkapkan model NLP yang paling efektif untuk analisis sentimen.

**Kata Kunci:** Tweet, Analisis Sentimen, NLP, SVM, Regresi Logistik, Random Forest, Neural Network, Naive Bayes

### Abstract:

The technological sector periodically undergoes fluctuations, and the term "Tech Winter" has been coined to describe periods of downturn. This study aims to analyze public sentiment towards Tech Winter on Twitter, leveraging the capabilities of Natural Language Processing (NLP). Employing a range of NLP models—Logistic Regression, Support Vector Machine (SVM), Random Forest, Neural Network, and Naive Bayes—this research compares their effectiveness in deciphering the sentiments expressed in tweets. The findings offer an understanding of public opinion during this period and shed light on the most effective NLP models for sentiment analysis.

**Keyword:** Tweet, Sentiment Analysis, NLP, SVM, Logistic Regression, Random Forest, Neural Network, Naive Bayes

---

## PENDAHULUAN

---

Pertukaran informasi dewasa ini mengalami kemudahan disebabkan kemajuan teknologi dan keberadaan sosial media. Sosial media menjadi alat masyarakat untuk menyebarluaskan dan mendapatkan informasi serta mengemukakan opini. Dari berbagai macam sosial media, twitter merupakan salah satu sosial media yang banyak digunakan (Anasati et al., 2015). Twitter merupakan platform sosial media yang memungkinkan penggunanya untuk membuat unggahan dan berinteraksi melalui tulisan yang disebut dengan tweet. Tweet pada awalnya dibatasi maksimal 140 karakter tetapi

kemudian diperluas menjadi 280 karakter. Penyebaran informasi melalui twitter bersifat real-time dan memiliki fitur trending topic apabila topik-topik yang muncul banyak dibahas penggunanya. Salah satu topik yang cukup banyak dibicarakan di twitter adalah fenomena tech winter.

Tech winter merupakan istilah untuk menggambarkan keadaan dimana banyak startup yang umumnya berbasis teknologi mulai berguguran (Linggar, 2022). Periode tech winter ditandai dengan pemutusan hubungan karyawan maupun hiring freeze atau menghentikan sementara rekrutmennya dalam rangka efisiensi. Selama periode ini, startup melakukan perubahan dan inovasi agar dapat beradaptasi, bertahan hingga akhirnya dapat tetap berkembang ketika sektor teknologi mengalami pemulihan (Breiman, 2001). Dengan banyaknya startup yang melakukan efisiensi, maka opini-opini dalam pembahasan seputar tech winter ini juga semakin banyak. Opini-opini tersebut dapat dianalisa dengan menggunakan analisis sentimen.

Analisis sentimen merupakan metode analisa berbasis teks di media sosial maupun platform internet lainnya sehingga dapat menjadi sumber data (Hintn, 1992). Analisis sentimen dilakukan untuk melihat polaritas opini terhadap sebuah objek oleh seseorang, apakah beropini positif atau negatif. Pada penelitian ini, metode analisis sentimen menggunakan Natural Language Processing.

Natural Language Processing dapat memecahkan masalah untuk memahami bahasa alami manusia dengan segala aturan gramatika dan semantiknya, dan mengubah bahasa tersebut menjadi representasi formal yang dapat diproses oleh computer (Pustejovsky, 2012). Dataset yang digunakan dibatasi dari mulai bulan januari 2021.

Dalam studi Wibowo (2020), analisis sentimen Twitter tentang kesehatan mental selama pandemi COVID-19 di Indonesia dilakukan dengan memanfaatkan data bertopik #dirumahaja dari 10 daerah. Menggunakan metode web scraping dengan library "GetOldTweets3", penelitian ini menghasilkan visualisasi yang komprehensif. Visualisasi ini mencakup analisis jumlah tweet, distribusi sentimen, dan kata-kata kunci populer, memberikan wawasan mendalam tentang sentimen publik terkait kesehatan mental selama pandemi (Dewandaru et al., 2022).

Dalam penelitian Shobrina and Chaerur (2022), analisis sentimen terhadap terorisme di Twitter dilakukan menggunakan teknik Natural Language Processing (NLP). Data diperoleh dari Kaggle, dan fokus penelitian ini adalah mengevaluasi pengaruh penggunaan Long Short Term Memory (LSTM) dalam NLP untuk menganalisis sentimen masyarakat terhadap terorisme (Callum et al., 1998). Hasil penelitian menunjukkan bahwa penggunaan LSTM dengan nilai unit sebesar 20 menghasilkan akurasi tertinggi sebesar 59.34%, yang lebih baik dibandingkan dengan nilai unit lainnya seperti 10, 30, hingga 100. Penelitian ini memberikan wawasan mengenai efektivitas LSTM dalam analisis sentimen terhadap isu terorisme di media sosial (Fathoniah et al., 2022).

Dalam studi Muktafin, Kusri, dan Luthfi (2020), penelitian dilakukan pada analisis sentimen ulasan produk di marketplace Shopee menggunakan pendekatan Natural Language Processing (NLP). Fokus penelitian ini adalah meningkatkan akurasi klasifikasi melalui pra-pemrosesan data, termasuk penggunaan fitur Word Normalizer untuk mengoreksi bahasa tidak baku dan slang, serta fitur Stemming dan Stopword removal (Kohavi, 1995). Hasil penelitian menunjukkan bahwa penerapan fitur NLP ini berhasil meningkatkan nilai akurasi menjadi 76,92%, presisi 80,00%, dan recall 74,07%. Ini merupakan peningkatan yang signifikan dibandingkan dengan klasifikasi tanpa fitur NLP yang hanya mencapai akurasi 69,23%, presisi 80,00%, dan recall 64,52%. Studi ini menunjukkan teknik pra-pemrosesan dalam analisis sentimen NLP menjadi penting, terutama dalam konteks ulasan produk di marketplace (Muktafin et al, 2020).

Dalam penelitian Munasatya and Novianto (2020), Natural Language Processing (NLP) diterapkan untuk menganalisis sentimen terhadap Presiden Jokowi menggunakan metode Multi Layer Perceptron.

Studi ini menggunakan dataset final yang berisi 3686 data dengan field teks atau opini Twitter, di mana label positif diberi nilai 1 dan negatif -1. Hasil pengujian menunjukkan bahwa jumlah hidden layer, learning rate, shuffling, dan dropout berpengaruh terhadap akurasi yang dihasilkan. Skenario terbaik yaitu skenario 7, menggunakan dataset Presiden Jokowi dengan 128 hidden layer dan tanpa pengacakan data, menghasilkan akurasi terbaik pada test accuracy sebesar 93,26%, f-measure 95,36%, dan nilai recall 94.21%. Penelitian ini menunjukkan efektivitas metode Multi Layer Perceptron dalam analisis sentimen terkait tokoh politik di media sosial (Munasatya et al., 2020).

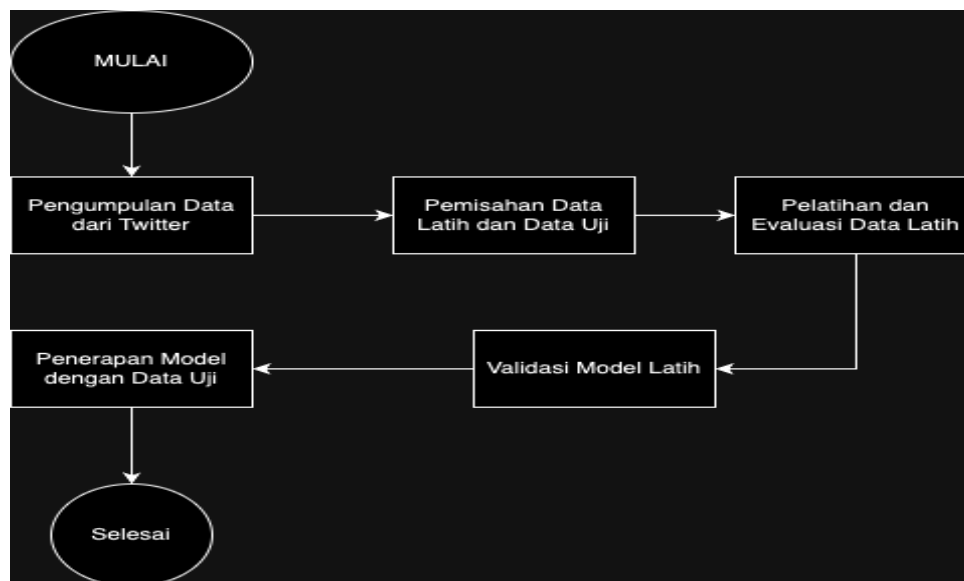
Dalam penelitian Gunawan, Pratiwi, dan Pratama (2018), analisis sentimen dilakukan pada ulasan produk dengan menggunakan metode Naïve Bayes Classifier. Data terkumpul dari 300 ulasan per kelas, yang diambil dari ulasan female. Pembobotan menggunakan metode TF-IDF, dan klasifikasi dilakukan dengan Naïve Bayes. Pengujian model melibatkan penggunaan confusion matrix, serta perhitungan akurasi, presisi, dan recall. Hasil pengujian menunjukkan bahwa dengan dataset yang terdiri dari 80% data latih dan 20% data uji, akurasi terbaik yang dicapai adalah 73,89%. Penelitian ini memberikan wawasan tentang efektivitas metode Naïve Bayes dalam analisis sentimen pada ulasan produk (Gunawan et al., 2018).

## METODE PENELITIAN

### Skema Alur Penelitian

Analisis sentimen merupakan topik studi yang menghususkan diri dalam mengevaluasi dan menginterpretasikan pendapat, emosi, dan sikap individu terhadap beragam subjek. Subjek-subjek ini dapat mencakup produk, layanan, orang-orang, kegiatan, atau topik-topik tertentu (Agarwal, 2018). Pendekatan analitis ini memegang peranan kunci dalam menentukan apakah pandangan masyarakat terhadap suatu isu cenderung positif atau negatif. Pemahaman ini sangat vital dalam upaya peningkatan kualitas produk atau layanan yang ditawarkan (Zhang et al., 2012).

Digambarkan sebagaimana gambar 1 berikut ini.



Gambar 1. Skema Alur Penelitian

## Pengumpulan Data

Data untuk analisis sentimen diperoleh secara manual dari platform Twitter dan berjumlah 161 data. Kumpulan data ini mencakup tweet yang diposting mulai dari tanggal 1 Januari 2022, dengan pencarian difokuskan pada tweet yang mengandung kata 'Tech Winter' dan 'layoff'. Untuk memastikan data yang relevan dengan konteks di Indonesia, pencarian dilakukan dengan filter bahasa, sehingga hanya tweet dalam Bahasa Indonesia yang diambil. Proses pengumpulan data manual ini mengharuskan peneliti untuk mencari, memilih, dan mengunduh tweet yang relevan untuk disimpan dalam format yang dapat dianalisis lebih lanjut.

## Pra-pemrosesan Data

Pra-pemrosesan data dilakukan dengan mengimplementasikan langkah-langkah sebagai berikut:

- Konversi teks ke huruf kecil.
- Penggantian kata informal dengan bentuk baku menggunakan kamus penggantian yang telah ditentukan.
- Penghapusan URL, karakter non-alfanumerik, dan emotikon menggunakan ekspresi reguler.
- Eliminasi stopword dan stemming untuk Bahasa Indonesia menggunakan library Sastrawi.
- Tokenisasi teks menjadi kata-kata individu.

## Ekstraksi Fitur dan Seleksi

Teks yang sudah di proses akan di ekstraksi menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Selanjutnya, akan dilakukan seleksi fitur menggunakan tes chi-kuadrat untuk mengidentifikasi 150 fitur teratas yang paling relevan untuk analisis sentimen.

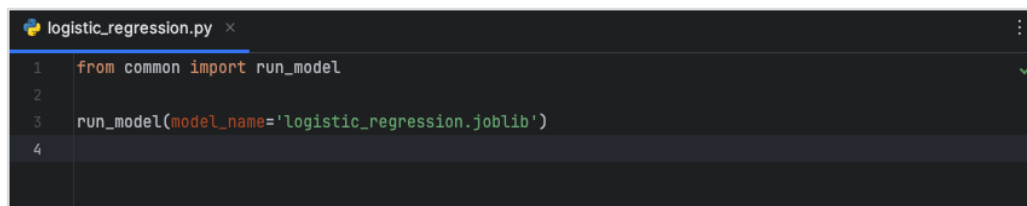
## Pelatihan dan Evaluasi Model

Dalam penelitian ini beberapa model akan diuji, diantaranya Regresi Logistik, SVM (Support Vector Machine), Random Forest, Neural Network (Multi-layer Perceptron), dan Naive Bayes. Model-model ini dilatih menggunakan dataset yang telah dibagi menjadi set pelatihan dan pengujian dengan proporsi 80:20. Model dievaluasi berdasarkan akurasi dalam mengklasifikasikan sentimen sebagai positif, atau negatif. Laporan klasifikasi detail dihasilkan untuk setiap model agar peneliti bisa menilai performa klasifikasi sentimen.

## Validasi Model

Validasi model dilakukan menggunakan teknik cross-validation dengan lipatan sebanyak 5 kali untuk memastikan kestabilan dan keumuman model. Hasil cross-validation mencakup akurasi rata-rata dan deviasi standar dari akurasi, yang memberikan gambaran mengenai konsistensi performa model yang sudah dilatih.

## Penerapan



```
logistic_regression.py ×  
1 from common import run_model  
2  
3 run_model(model_name='logistic_regression.joblib')  
4
```

```
10 usages
def run_model(model_name='logistic_regression.joblib'):
    model = load(model_name)

    # Load the saved TF-IDF vectorizer
    tfidf_vectorizer = load('tfidf_vectorizer.joblib')

    # Load the saved k_best selector
    k_best = load('k_best_selector.joblib')

    # Load the data without sentiment
    data_without_sentiment = pd.read_csv('data/data-without-sentiment.csv')

    data_without_sentiment['processed_tweet'] = data_without_sentiment['tweet'].apply(preprocess_text)

    x = tfidf_vectorizer.transform(data_without_sentiment['processed_tweet'])
    x = k_best.transform(x)

    # Predict sentiment
    predictions = model.predict(x)
    data_without_sentiment['predicted_sentiment'] = predictions

    # Save the dataframe with predictions to a new CSV file
    output_filename = model_name.replace(__old: '.joblib', __new: '_predictions.csv')
    data_without_sentiment.to_csv(output_filename, index=False)

    # Generate a report
    sentiment_counts = data_without_sentiment['predicted_sentiment'].value_counts()

    print("Sentiment Analysis Report:")
    print(sentiment_counts)
```

Model yang sudah dilatih akan disimpan ke disk menggunakan joblib, yang memungkinkan penggunaan model tersebut dalam prediksi sentimen pada data uji ataupun data baru dimasa depan. Fungsi run\_model diatas mendemonstrasikan bagaimana model dapat dioperasikan untuk memuat dan memprediksi sentimen dari tweet yang belum dilabeli.

## HASIL DAN PEMBAHASAN

Penelitian ini berhasil mengimplementasikan dan mengevaluasi berbagai model pembelajaran mesin untuk analisis sentimen tweet yang berkaitan dengan 'Tech Winter' dan 'layoff' dalam Bahasa Indonesia. Berikut adalah hasil dan pembahasan dari masing-masing model:

### Logistic Regression

Model Regresi Logistik menunjukkan akurasi sebesar 83,33% dengan presisi yang tinggi untuk kelas negatif. Hal ini menandakan kemampuan model untuk mengidentifikasi tweet negatif dengan sangat baik. Namun, model ini memiliki recall yang lebih rendah untuk kelas positif, yang mengindikasikan bahwa beberapa tweet positif tidak terdeteksi dengan baik (Turban, 2005). F1-score untuk kelas positif adalah 0.73, menunjukkan ruang untuk perbaikan dalam keseimbangan antara presisi dan recall.

Tabel 1. Model Logistic Regression

	Precision	Recall	F1-Score	Score
Negative	0.79	1.00	0.88	11
Positive	1.00	0.57	0.73	7
Accuracy			0.83	18
Macro Average	0.89	0.79	0.80	18
Weighted Average	0.87	0.83	0.82	18

### Support Vector Machine (SVM)

Model SVM mencapai akurasi yang sangat tinggi yaitu 94,44%, dengan presisi dan recall yang luar biasa untuk kedua kelas sentimen. F1-score yang tinggi menunjukkan bahwa model ini efektif dalam klasifikasi sentimen secara keseluruhan dan merupakan kandidat terbaik berdasarkan pengukuran ini (Lestari et al., 2017).

**Tabel 2. Model SVM**

	Precision	Recall	F1-Score	Score
Negative	0.92	1.00	0.96	11
Positive	1.00	0.86	0.92	7
Accuracy			0.94	18
Macro Average	0.96	0.93	0.94	18
Weighted Average	0.95	0.94	0.94	18

### Random Forest

Random Forest menunjukkan akurasi yang lebih rendah dibandingkan dengan model SVM, yaitu 77,78%. Meskipun presisi dan recall cukup seimbang untuk kedua kelas, model ini masih memiliki potensi untuk ditingkatkan, terutama dalam meningkatkan sensitivitas terhadap kelas positif (Rusdianan et al., 2019).

**Tabel 3. Random Forest**

	Precision	Recall	F1-Score	Score
Negative	0.82	0.82	0.82	11
Positive	0.71	0.71	0.71	7
Accuracy			0.78	18
Macro Average	0.77	0.77	0.77	18
Weighted Average	0.78	0.78	0.78	18

### Neural Network

Neural Network (Jaringan Saraf Tiruan) menunjukkan performa yang kuat dengan akurasi 88,89%. Model ini menunjukkan presisi yang baik untuk kelas negatif dan sempurna untuk kelas positif, namun masih memerlukan peningkatan dalam hal recall untuk kelas positif (Hosmer et al., 2019).

**Tabel 4. Neural Network**

	Precision	Recall	F1-Score	Score
Negative	0.85	1.00	0.92	11
Positive	1.00	0.71	0.83	7
Accuracy			0.89	18
Macro Average	0.92	0.86	0.88	18
Weighted Average	0.91	0.89	0.88	18

### Naive Bayes

Model Naive Bayes juga menunjukkan akurasi yang tinggi, sebanding dengan SVM, yaitu 94,44%. Dengan presisi dan recall yang tinggi untuk kedua kelas, model ini menunjukkan kinerja yang sangat baik dalam klasifikasi sentimen (Cortes et al., 1995).

**Tabel 5. Naive Bayes**

	Precision	Recall	F1-Score	Score
Negative	0.92	1.00	0.96	11
Positive	1.00	0.86	0.92	7
Accuracy			0.94	18
Macro Average	0.96	0.93	0.94	18
Weighted Average	0.95	0.94	0.94	18

### Cross-Validation

Hasil cross-validation untuk model-model di atas menunjukkan bahwa Neural Network memiliki mean accuracy tertinggi sebesar 95,56% dengan standar deviasi terendah, menandakan konsistensi performa yang sangat baik. Di sisi lain, Random Forest memiliki mean accuracy terendah yaitu 72,22%.

**Tabel 6. Random Forest**

Model	Mean Accuracy	Standard Deviation
Logistic Regression	0.7444	0.1144
SVM	0.7778	0.0786
Random Forest	0.7222	0.093
Neural Network	0.9556	0.0416
Naive Bayes	0.8444	0.0737

### KESIMPULAN

Berdasarkan hasil yang diperoleh, dapat dilihat bahwa SVM dan Naive Bayes memberikan hasil yang sangat baik dalam hal akurasi dan metrik evaluasi lainnya. Namun, Neural Network menunjukkan keunggulan dalam validasi silang, yang menunjukkan konsistensi performanya yang tinggi dan kemungkinan kinerja yang lebih baik pada data yang belum dilihat. Selanjutnya, penelitian ini juga mengidentifikasi beberapa area untuk peningkatan, termasuk peningkatan recall untuk kelas positif dalam model Regresi Logistik dan Random Forest.

Berikut merupakan hasil pelabelan data dari 71 data yang belum dilabeli:

Model	Positive	Negative
Logistic Regression	66	5
SVM	58	13
Random Forest	48	23
Neural Network	56	15
Naive Bayes	62	9
Model	Positive	Negative

## BIBLIOGRAPHY

---

- Anasari, N., & Handoyo, P. (2015). Media Sosial Sebagai Panggung Drama ( Studi Deskriptif Presentasi Diri Pengguna Twitter di Kalangan Mahasiswa Unesa ), *Paradigma*, 3(3)
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. DOI:10.1023/A:1010933404324
- C. Aggarwal (2018). "Opinion Mining and Sentiment Analysis," *Data Mining and Knowledge Discovery Series*. DOI: 10.1007/978-3-319-73531-3\_13.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297. DOI: 10.1007/BF00994018
- Dewandaru, A., & Wibowo, J. S. (2022). Analisis Sentimen dan Klasifikasi Tweet Terkait Mutasi COVID-19 menggunakan Metode Naïve Bayes Classifier. *Jurnal Teknologi dan Manajemen Informatika*, 8(1), 32-38.
- Fathoniah, S., & Rozikin, C. (2022). Analisis Sentimen Masyarakat terhadap Teroris dalam Media Sosial Twitter menggunakan NLP. *Jurnal Ilmiah Wahana Pendidikan*, 8(13), 412-419. DOI: 10.5281/zenodo.6962682
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sentiment Analysis System In Product Reviews Using Naive Bayes Method. *J. Education And Research. Inform*, 4(2), 113. DOI: 10.26418/jp.v4i2.27526
- Hinton, G. E. (1992). How Neural Networks Learn from Experience. *Scientific American*, 267(3), 144–151. doi: 10.1038/scientificamerican0992-144.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, 14(2), 1137-1145. DOI: 10.5555/1643031
- L. Zhang and B. Liu (2012). "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers. p. 1-167. DOI: 10.2200/s00416ed1v01y201204hlt016.
- Lestari, A. R. T., Perdana, R. S., & Fauzi, M. A. (2017). Analisis sentimen tentang opini pilkada dki 2017 pada dokumen twitter berbahasa indonesia menggunakan naive bayes dan pembobotan emoji. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), 1718-1724.
- Linggar, (2022). 3 Tips Startup Bertahan Melewati Tech Winter
- McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, 752(1), 41-48. DOI: 10.3115/1067807.1067848
- Muktafin, E. H., Kusri, K., & Luthfi, E. T. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Jurnal Eksplorasi Informatika*, 10(1), 32-42. DOI: 10.30864/eksplorasi.v10i1.390
- Munasatya, N., & Novianto, S. (2020). Natural Language Processing untuk Sentimen Analisis Presiden Jokowi Menggunakan Multi Layer Perceptron. *Techno. Com*, 19(3), 237-244. DOI: 10.33633/tc.v19i3.3630
- Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. " O'Reilly Media, Inc."
- Rusdian, D., & Rosiyadi, D. (2019). Analisa Sentimen terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), 230-235. DOI: 10.24114/cess.v4i2.13796
- Turban, E. (2005). *Decision Support System and Intelligent System (Sistem Pendukung Keputusan dan Sistem Cerdas)*. Yogyakarta: Andi Offset



**Copyright holder:**

Rifky Aziz Fauzianto<sup>1</sup>, Supatman<sup>2</sup> (2023)

**First publication right:**

Jurnal Syntax Admiration

**This article is licensed under:**

