

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: <https://doi.org/10.35870/jtik.v9i1.3080>

Analisis Sentimen terhadap Perpanjangan Masa Jabatan Presiden Indonesia Menggunakan *Naïve Bayes*

Zuhdi Hanif¹, Untung Surapati^{2*}

^{1,2*} Program Studi Teknik Informatika, Fakultas Teknologi Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

article info

Article history:

Received 7 August 2024

Received in revised form

15 September 2024

Accepted 25 October 2024

Available online January 2025.

Keywords:

Sentiment Analysis; Naïve Bayes; President of Indonesia; Sentiment Score; Twitter.

Kata Kunci:

Analisis Sentimen; Naïve Bayes; Presiden Indonesia; Sentimen Skor; Twitter.

abstract

Twitter social media is often used as a medium to express opinions on the government in Indonesia. There are also many controversial things on Twitter social media against the government, as is the case today there is a controversy whether there is a proposal to make the term of office of the President in Indonesia three periods, which previously could only serve up to two terms or ten years for one President. Public opinion or sentiment which in Twitter's terms is commonly referred to as "shrink" can be in the form of negative or positive opinions. However, the amount of data is quite large so it takes a method that can be used to make it happen, namely by using sentiment analysis. Sentiment analysis can be used as a solution to process these opinions using the Naïve Bayes Classifier algorithm. The results of the Naïve Bayes technique require evaluation to determine the best model. The test is carried out with three models, namely 70:30, 80:20, and 90:10 which are then evaluated using a confusion matrix. Based on the comparison of the evaluation test results, the best scenario for the Naïve Bayes classification model is in the first scenario or model (90% training data and 10% testing data) with an accuracy value of 95%, a precision value of 97%, a recall value of 96%, and the f-measure value is 96%.

abstrak

Media sosial twitter sering digunakan sebagai media untuk menyampaikan pendapat terhadap pemerintahan di Indonesia. Banyak juga hal-hal yang kontroversial di media sosial twitter terhadap pemerintah, seperti halnya pada hari ini sedang terjadi kontroversi dimana ada sebuah usulan untuk menjadikan masa jabatan Presiden di Indonesia yaitu tiga periode, yang dimana tadinya hanya bisa menjabat sampai dua periode atau sepuluh tahun saja untuk satu orang Presiden. Opini atau sentimen masyarakat yang dalam istilah Twitter biasa disebut sebagai "ciutan" ini dapat berupa opini negatif maupun positif. Akan tetapi jumlah data tersebut cukup banyak sehingga dibutuhkan suatu metode untuk mewujudkannya yaitu dengan menggunakan analisis sentimen. Analisis sentimen dapat dijadikan solusi untuk mengolah opini tersebut dengan menggunakan algoritma Naïve Bayes Classifier. Pada hasil dari teknik Naïve Bayes memerlukan evaluasi untuk menentukan model terbaik. Pengujian dilakukan dengan tiga model yaitu 70:30, 80:20, dan 90:10 yang kemudian dievaluasi menggunakan confusion matrix. Berdasarkan perbandingan hasil pengujian evaluasi, skenario terbaik yang untuk model klasifikasi Naïve Bayes yaitu pada skenario atau model pertama (90% data training dan 10% data testing) dengan nilai accuracy sebesar 95%, nilai precision sebesar 97%, nilai recall sebesar 96%, dan nilai f-measure sebesar 96%.

Corresponding Author. Email: untungsurapati@gmail.com ^{2}.



ACM Computing Classification System (CCS)

EBSCOhost

Communication and Mass Media Complete (CMMC)

Copyright 2025 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISEI). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

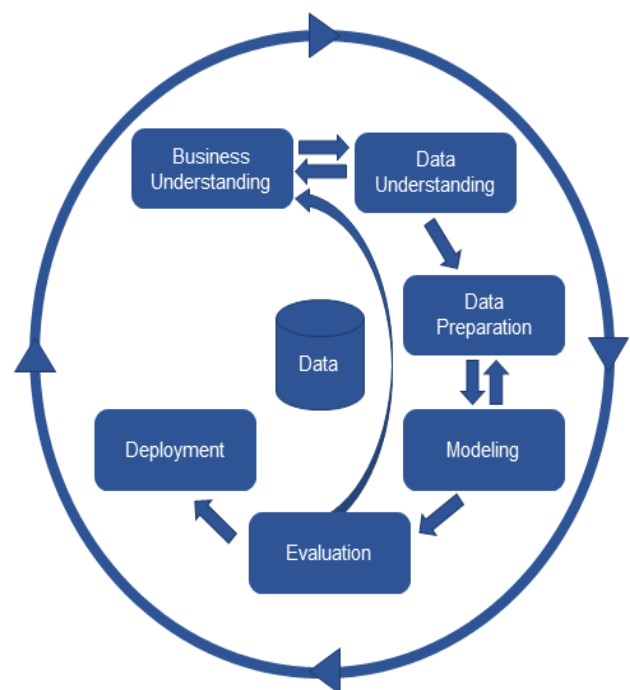
1. Pendahuluan

Media sosial *Twitter* sering digunakan sebagai media untuk menyampaikan pendapat terhadap pemerintahan di Indonesia. Media ini juga sering dipakai oleh pemerintah untuk mensosialisasikan berbagai informasi terkait pemerintahan Indonesia sehingga masyarakat dapat mengetahui informasi tersebut. Banyak hal kontroversial di media sosial *Twitter* yang berkaitan dengan pemerintah, misalnya saat ini sedang muncul kontroversi terkait usulan perpanjangan masa jabatan Presiden di Indonesia menjadi tiga periode, yang mana sebelumnya hanya dapat menjabat hingga dua periode atau sepuluh tahun untuk satu Presiden. Wacana tersebut menjadi kontroversi karena beberapa survei menunjukkan ketidaksetujuan dari masyarakat terhadap wacana tersebut. Salah satunya adalah survei oleh lembaga *Indonesia Political Opinion* (IPO) yang menunjukkan bahwa 61% responden tidak setuju, sementara 39% responden setuju (Medistiara, 2022).

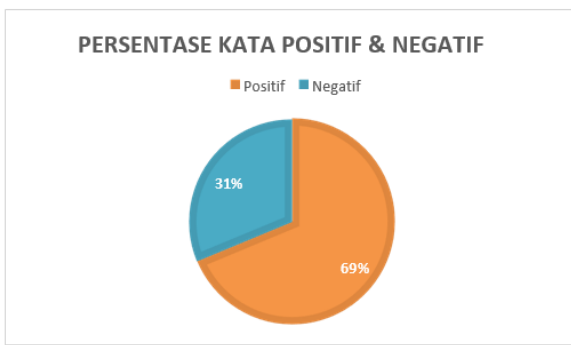
Namun, Menteri Koordinator Bidang Kemaritiman dan Investasi, Luhut B. Pandjaitan, menyatakan bahwa ia memiliki *big data* yang menunjukkan bahwa mayoritas rakyat Indonesia mendukung penundaan pemilu 2024 dan perpanjangan masa jabatan Presiden. Hal ini ia sampaikan dalam podcast *Close the Door* bersama Deddy Corbuzier (Santoso & Fadil Djailaini, 2022). Opini atau sentimen masyarakat yang dalam istilah *Twitter* biasa disebut sebagai “ciutan” dapat berupa opini negatif maupun positif. Akan tetapi, jumlah data yang sangat besar ini memerlukan metode khusus untuk pengolahannya, yaitu dengan menggunakan analisis sentimen. Analisis sentimen adalah teknik pembersihan data dari kata dan simbol yang tidak relevan, serta mengubah data yang bersifat kualitatif menjadi data kuantitatif. Kemudian, data ulasan pengguna diklasifikasikan untuk mendapatkan ulasan yang bersifat positif atau negatif (Manik, Ernawati, & Nurlaili, 2021). Saat ini, analisis sentimen merupakan salah satu topik yang banyak digunakan oleh para peneliti, yang bertujuan untuk menyediakan informasi dari sebuah dataset yang tidak terstruktur. Menurut Arsi & Waluyo (2021), analisis sentimen ini juga dapat diterapkan pada opini di berbagai bidang seperti ekonomi, politik, sosial, dan hukum. Analisis sentimen juga mampu mengelompokkan polaritas.

2. Metodologi Penelitian

Systematic Literature Review (SLR) adalah proses sistematis yang menyeluruh untuk mengidentifikasi, mengevaluasi, dan menganalisis penelitian yang relevan dengan topik tertentu. SLR bertujuan untuk memberikan gambaran yang jelas tentang penelitian yang telah dilakukan dalam suatu bidang ilmu. Metode SLR dianggap sebagai metode tingkat tinggi dalam menghasilkan pemahaman objektif terhadap suatu topik, karena melibatkan proses sistematis, kritis, dan menyeluruh dalam mengevaluasi dan menyintesis bukti-bukti ilmiah yang ada. Pengklasifikasi *Naïve Bayes* menggunakan probabilitas bersyarat untuk setiap kelas yang mungkin, diturunkan dari matriks data, dan didasarkan pada *teorema Bayes*. Asumsi dasar *Naïve Bayes Classifier* (NBC) mengandaikan bahwa variabel-variabel sistem bersifat independen satu sama lain. Ketika diterapkan pada masalah klasifikasi dalam dunia nyata, pengklasifikasi berbasis NBC tetap menghasilkan hasil yang dapat diterima, meskipun bergantung pada asumsi yang sering kali tidak sepenuhnya valid (Marcos de Moraes, Soares, & Machado, 2020).



Gambar 1. Tahapan Crisp-DM



Gambar 2. Persentase Kata Positif & Negatif



Gambar 3. *WordCloud*

3. Hasil dan Pembahasan

Hasil

Penelitian ini menggunakan dataset dari *tweet* dengan kata kunci "jokowi tiga periode" melalui teknik *crawling data* pada *Twitter API*. Data yang terkumpul berjumlah 1.000 *tweet* dari tanggal 1 Juni 2022 hingga 30 Juni 2022. Setelah melalui tahap pembersihan, data yang digunakan menjadi 641 *tweet* yang terdiri dari 422 *tweet* positif dan 192 *tweet* negatif. Persentase sentimen positif dan negatif dapat dilihat pada gambar persentase sentimen. *Word cloud* pada data tersebut menunjukkan kata-kata yang sering muncul, seperti “jokowi”, “presiden”, dan “menteri.” Semakin besar ukuran kata pada *word cloud*, semakin tinggi frekuensi kemunculan kata tersebut, menunjukkan intensitas penggunaannya dalam cuitan di *Twitter*. Proses *data mining* dilakukan menggunakan algoritma *Naïve Bayes* dengan tiga skenario, yaitu 90% data *training* dan 10% data *testing*, 80% data *training* dan 20% data *testing*, serta 70% data *training* dan 30% data *testing*. Hasil performa dari tiap skenario dibandingkan untuk mendapatkan model terbaik.

Berdasarkan hasil evaluasi, skenario terbaik untuk model klasifikasi *Naïve Bayes* terdapat pada skenario pertama (90% data *training* dan 10% data *testing*) dengan nilai akurasi sebesar 95%, *precision* sebesar 97%, *recall* sebesar 96%, dan *f-measure* sebesar 96%. Model ini dinilai optimal karena memperoleh persentase nilai tertinggi pada tiga indikator pengujian yaitu akurasi, *precision*, dan *f-measure*, serta nilai *recall* yang tetap baik meskipun bukan yang tertinggi di antara skenario lainnya. Penelitian ini juga melakukan analisis sentimen terhadap data komentar di *Instagram* terkait usulan perpanjangan masa jabatan kepala desa. Data komentar diklasifikasikan ke dalam tiga kelas, yaitu positif, negatif, dan netral, dengan menggunakan algoritma *Naïve Bayes Classifier*. Evaluasi sistem dilakukan menggunakan *Confusion Matrix* untuk mengetahui nilai akurasi, *precision*, *recall*, dan *f-measure* dari model tersebut. Data yang diperoleh dari proses pengambilan (*crawling*) masih dalam kondisi yang kurang terstruktur dan mengandung berbagai gangguan seperti tanda baca, angka, simbol, serta kata-kata yang tidak resmi. Elemen-elemen ini dianggap tidak relevan dalam proses pengelompokan. Oleh karena itu, langkah selanjutnya adalah melakukan pra-pemrosesan pada dataset dalam tahap *pre-processing*. Penelitian juga mencakup data dari *Twitter* dengan menggunakan kata kunci pencarian seperti #calonpresiden, #prabowo, #anies, dan #ganjar. Data tersebut adalah *tweet* berbahasa Indonesia yang diunduh dari *Twitter* dengan total 3.862 *tweet* yang kemudian disimpan dalam format *CSV* di database.

Pembahasan

Penelitian ini mengkaji sentimen masyarakat terkait usulan perpanjangan masa jabatan Presiden Indonesia melalui analisis data *tweet* di media sosial *Twitter* menggunakan algoritma *Naïve Bayes*. Hasil analisis menunjukkan bahwa opini masyarakat terbagi, dengan proporsi sentimen positif yang lebih tinggi dibandingkan negatif. Hal ini sejalan dengan penelitian Appel *et al.* (2020) yang menyoroti peran media sosial sebagai platform untuk mengungkapkan pandangan publik terhadap isu-isu kebijakan, di mana berbagai opini dapat tercermin dengan cepat melalui peningkatan volume percakapan daring. Penggunaan algoritma *Naïve Bayes* dalam penelitian ini menunjukkan efektivitasnya untuk klasifikasi sentimen dengan akurasi yang cukup tinggi pada skenario terbaik (90% data *training* dan 10% data

testing) yang menghasilkan nilai akurasi sebesar 95%, *precision* 97%, *recall* 96%, dan *f-measure* 96%. Marcos de Moraes *et al.* (2020) mendukung temuan ini dengan menunjukkan bahwa *Naïve Bayes* mampu menghasilkan performa klasifikasi yang andal meskipun pada data yang tidak terstruktur. Dalam konteks analisis sentimen media sosial, ketahanan algoritma ini terhadap kompleksitas data menjadi faktor penting untuk mempertahankan akurasi klasifikasi.

Proses *pre-processing* yang mencakup pembersihan data dari elemen yang tidak relevan seperti tanda baca, angka, simbol, dan kata-kata tidak resmi sangat berperan dalam meningkatkan performa model, sebagaimana dinyatakan oleh Badjrie *et al.* (2021), bahwa pembersihan dan normalisasi data sangat penting dalam proses analisis sentimen untuk mengurangi bias pada hasil klasifikasi. Langkah ini memberikan kejelasan pada data, memungkinkan algoritma untuk fokus pada konten yang bermakna dan berhubungan langsung dengan topik sentimen yang dianalisis. Sebagai tambahan, *word cloud* yang digunakan dalam penelitian ini menunjukkan kata-kata yang sering muncul seperti "jokowi," "presiden," dan "menteri," yang mengindikasikan bahwa isu kepemimpinan nasional menjadi fokus utama dalam percakapan publik di *Twitter*. Aulia & Patriya (2019) mengemukakan bahwa visualisasi *word cloud* dapat membantu untuk memahami intensitas dan fokus sentimen publik dalam data yang besar, terutama pada topik yang sensitif secara politik.

Penelitian ini juga berkontribusi terhadap literatur yang semakin berkembang mengenai analisis sentimen dalam konteks isu kebijakan publik di media sosial. Sebagai contoh, Arsi & Waluyo (2021) dan Manik *et al.* (2021) telah menunjukkan bagaimana analisis sentimen dengan berbagai algoritma, seperti *Support Vector Machine* (SVM) dan *Convolutional Neural Network* (CNN), dapat mengidentifikasi respons publik terhadap kebijakan pemerintah, dan hasilnya dapat membantu dalam memahami kecenderungan opini serta persepsi masyarakat pada isu-isu terkini. Secara keseluruhan, penelitian ini memberikan wawasan penting mengenai sentimen masyarakat terhadap isu perpanjangan masa jabatan Presiden melalui media sosial, yang dapat berfungsi sebagai referensi bagi pemerintah dan pengambil kebijakan

untuk merespons persepsi publik secara lebih terinformasi. Mengingat bahwa media sosial seperti *Twitter* menyediakan sumber data yang kaya dan langsung dari masyarakat, analisis yang lebih komprehensif menggunakan algoritma yang berbeda, seperti gabungan *Naïve Bayes* dengan metode berbasis jaringan saraf tiruan atau *machine learning*, dapat menjadi arah penelitian selanjutnya untuk mengatasi kompleksitas yang lebih tinggi dan meningkatkan keakuratan klasifikasi (Nida, 2020; Paulina *et al.*, 2020).

4. Kesimpulan

Penelitian ini menganalisis sentimen masyarakat di *Twitter* terkait usulan perpanjangan masa jabatan Presiden Indonesia dengan menggunakan algoritma *Naïve Bayes*. Metode yang diterapkan dalam penelitian ini adalah *Knowledge Discovery in Database* (KDD), yang melibatkan lima tahapan utama. Tahap pertama adalah *data selection*, yang mencakup proses *crawling* data untuk mengumpulkan data mentah dari *Twitter*. Tahap kedua adalah *pre-processing*, yang terdiri dari lima langkah, yaitu *case folding*, *tokenizing*, *spelling normalization*, *filtering*, dan *stemming*, bertujuan untuk membersihkan dan mempersiapkan data agar lebih terstruktur. Selanjutnya, tahap *transformation* dilakukan dengan pembobotan kata menggunakan metode *TF-IDF* untuk menyoroti kata-kata yang paling relevan dalam dataset. Tahap keempat adalah *data mining*, di mana data diklasifikasikan menggunakan algoritma *Multinomial Naïve Bayes* untuk mengidentifikasi polaritas sentimen. Terakhir, tahap *evaluation* dilakukan dengan menggunakan *confusion matrix* untuk menghitung akurasi, *precision*, *recall*, dan *f-measure* dari model yang telah dibuat. Model ini berguna untuk mengklasifikasikan opini publik, termasuk kritik terhadap kepala pemerintahan yang disampaikan di *Twitter*, di mana interaksi di media sosial memiliki dampak yang beragam, tergantung pada intensitas penggunaan dan konten yang diakses.

5. Ucapan Terima Kasih

Dengan Penuh Rasa Syukur Dan Terima Kasih Kepada Keluarga Yang Selalu Memberikan Support Dan Saya Mahasiswa Dari Sekolah Tinggi Ilmu

Komputer Cipta Karya Informatika, Ingin Menyampaikan Ucapan Terima Kasih Yang Tulus Kepada Bapak Untung Surapati, SE.,SH.,MH.,MM.,M.KOM Atas Bimbingan, Arahkan, Dan Inspirasi Yang Luar Biasa Selama Pelaksanaan Skripsi Ini. Bapak Telah Memberikan Panduan Yang Sangat Berharga, Serta Kesabaran Dan Dukungannya Yang Tak Terhingga Selama Kami Menjalankan Tugas Ini. Saya Juga Ingin Menyampaikan Terima Kasih Kepada Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika Atas Fasilitas Dan Dukungan Yang Diberikan Selama Saya Menjalankan Skripsi Ini. Semua Bantuan Dan Dukungan Ini Sungguh Berarti Bagi Kelancaran Skripsi Saya. Dengan Rendah Hati, Kami Menyadari Bahwa terselesaikannya Skripsi Ini Tidak Mungkin Terwujud Tanpa Kontribusi Dari Semua.

6. Daftar Pustaka

- Ali, I., Asif, M., Hamid, I., Sarwar, M. U., Khan, F. A., & Ghadi, Y. (2022). A word embedding technique for sentiment analysis of social media to understand the relationship between Islamophobic incidents and media portrayal of Muslim communities. *PeerJ Computer Science*, 8, e838.
- Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing science*, 48(1), 79-95.
- Arsi, P., & Waluyo, R. (2021). Analisis sentimen wacana pemindahan ibu kota Indonesia menggunakan algoritma Support Vector Machine (SVM). *J. Teknol. Inf. dan Ilmu Komput*, 8(1), 147.
- Aulia, G. N., & Patriya, E. (2020). Implementasi Lexicon Based Dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019. *Jurnal Ilmiah Informatika Komputer*, 24(2), 140-153. DOI: <http://dx.doi.org/10.35760/ik.2019.v24i2.2369>.
- Badjrie, S. H., Pratiwi, O. N., & Anggana, H. D. (2021). Analisis Sentimen Review Customer Terhadap Produk Indihome Dan First Media Menggunakan Algoritma Convolutional Neural Network. *eProceedings of Engineering*, 8(5).
- Basit, A. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Hasil Panen Padi. *JTik (Jurnal Teknik Informatika Kaputama)*, 4(2), 208-213.
- Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the total environment*, 728, 138882.
- Hakimi, F. D. D. (2018). Sistem Analisis Sentimen Publik Tentang Opini Pemilihan Kepala Daerah Jawa Timur 2018 Pada Dokumen Twitter Menggunakan Naive Bayes Classifier. *Universitas Islam Negeri Sunan Ampel Surabaya*.
- Harun, A., & Ananda, D. P. (2021). Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naive bayes dan Decision Tree: Analysis of Public Opinion Sentiment About Covid-19 Vaccination in Indonesia Using Naive Bayes and Decision Tree. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 58-64. DOI: <https://doi.org/10.57152/malcom.v1i1.63>.
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.
- Lutz, M. (2001). *Programming python*. " O'Reilly Media, Inc."
- Maclean, F., Jones, D., Carin-Levy, G., & Hunter, H. (2013). Understanding twitter. *British journal of occupational Therapy*, 76(6), 295-298.
- Manik, G., Ernawati, I., & Nurlaili, I. (2021). Analisis Sentimen Pada Review Pengguna E-Commerce Bidang Pangan Menggunakan Metode Support Vector Machine (Studi Kasus: Review Sayurbox dan Tanihub pada Google Play). In *Prosiding*

Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya (Vol. 2, No. 2, pp. 64-74).

- Marcos de Moraes, R., Soares, E. A. D. M. G., & Machado, L. D. S. (2020). A double weighted fuzzy gamma naive bayes classifier. *Journal Of Intelligent & Fuzzy Systems*, 38(1), 577-588. DOI: <https://doi.org/10.31004/jpdk.v4i4.6199>.
- Merawati, D., & Rino, R. (2019). Penerapan Data Mining Penentu Minat Dan Bakat Siswa Smk Dengan Metode C4. 5. *ALGOR*, 1(1), 28-37.
- Nida, E. A. (2020). Analisis Kinerja Algoritma Support Vector Machine (SVM) Guna Pengambilan Keputusan Beli/Jual Pada Saham PT Elnusa Tbk. (ELSA). *Jurnal Transformatika*, 17(2), 160-170. DOI: <http://dx.doi.org/10.26623/transformatika.v17i2.1649>.
- Nugroho, R. A., Cholissodin, I., & Indriati, I. (2021). Implementasi Naïve Bayes Classifier untuk Klasifikasi Emosi Tweet Berbahasa Indonesia pada Spark. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(1), 301-310.
- Ratiasasadara, P. W., Sudarno, S., & Tarno, T. (2023). Analisis Sentimen Penerapan Ppkm Pada Twitter Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi-Square. *Jurnal Gaussian*, 11(4), 580-590.
- Sari, R. (2020). Analisis sentimen pada review objek wisata dunia fantasi menggunakan algoritma K-Nearest Neighbor (k-nn). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1). DOI: <https://doi.org/10.31294/evolusi.v8i1.7371>.
- Sihombing, R. E., Rachmatin, D., & Dahlan, J. A. (2019). Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan Metode K-Medoids Clustering. *Jurnal EurekaMatika*, 7(1), 58-79.
- Sumantri, R. B. B., & Utami, E. (2020). Penentuan Status Tahapan Keluarga Sejahtera Kecamatan Sidareja Menggunakan Teknik Data Mining. *Respati*, 15(3), 71-82.
- Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*, 8, 23522-23530.