

Analisis Sentimen Publik Terkait Bakal Calon Presiden 2024 Menggunakan Metode Support Vector Machine

Siti Aulia Rachmini*¹, Ismaun Rusman², Dian Megah Sari³, Arnita Irianti⁴, Supriadi Saparang⁵

^{1,2,3,4,5}Universitas Sulawesi Barat, Majene

e-mail: *¹ sitiaulia.rachmini@unsulbar.ac.id, ² maunkalmaliki@unsulbar.ac.id,
³ dianmegahsari@unsulbar.ac.id, ⁴ arnitairianti@unsulbar.ac.id,
⁵ supriadisaparang@gmail.com

Abstrak

Pemilihan Presiden tahun 2024 merupakan pemilihan setelah masa jabatan Presiden Joko Widodo selama 2 periode, periode terakhir dari Presiden Jokowi menuai banyak nama yang diperkirakan akan diusungkan ketika pilpres. Eleksabilitas bakal calon presiden dapat menjadi pandangan baru bagi masyarakat dalam menghadapi pilpres pada tahun 2024. Penelitian ini bertujuan untuk mensoroti tentang sentiment masyarakat yang ditujukan kepada bakal calon presiden 2024 dengan tujuh nama tersebut dengan mengevaluasi performa kandidat dalam pengujian atau pengukuran menggunakan metrik-metrik seperti accuracy, recall, dan precision. Penelitian ini melakukan analisis sentiment publik terkait bakal calon presiden 2024 menggunakan metode support vector machine dan didapatkan hasil bahwa akurasi tertinggi memiliki persentase 98,33% dan nilai akurasi terendah 76,67% sedangkan nilai recall tertinggi yaitu 0,90% dan nilai precision tertinggi 1,00% dimana nilai tersebut mewakili beberapa kandidat bakal calon presiden 2024.

Kata kunci : SVM, evaluasi, pengujian, pengukuran, *accuracy*, *recall*, *precision*

Abstract

The 2024 Presidential Election is an election after President Joko Widodo's 2-term tenure, the last period of President Jokowi reaped many names that are expected to be carried during the presidential election. The electability of presidential candidates can be a new view for the public in facing the presidential election in 2024. This study aims to highlight public sentiment aimed at 2024 presidential candidates with these seven names by evaluating candidate performance in testing or measurement using metrics such as accuracy, recall, and precision. This study conducted an analysis of public sentiment related to the 2024 presidential candidate using the support vector machine method and found that the highest accuracy had a percentage of 98.33% and the lowest accuracy value of 76.67% while the recall value was increased at 0.90% and the precision value was increased at 1.00% where the value represented several candidates for the 2024 presidential candidate.

Keywords : SVM, evaluation, testing, measurement, *accuracy*, *recall*, *precision*

1. PENDAHULUAN

Pemilihan presiden tahun 2024 mulai banyak diperbincangkan. Adapun nama-nama lain yang banyak diperbincangkan seperti Ganjar Pranowo, Ridwan Kamil, Prabowo Subianto, Sandiaga Uno, Agus Harimurti Yudhoyono dan Puan Maharani [1]. Sentimen masyarakat dalam twitter dapat dilihat dari

tweet atau teks yang digunakan untuk memberikan informasi di twitter, tweet biasa digunakan oleh pengguna internet untuk memberitahu tentang apa yang sedang dilakukan, percakapan, berbagi informasi, juga menuliskan atau mengekspresikan opini apa yang sedang dirasakan pengguna internet.

Fokus utama dalam penelitian ini adalah menyoroti tentang sentimen masyarakat yang ditujukan kepada bakal calon presiden 2024 dengan 7 nama yang sering muncul yaitu Anies Baswedan, Ganjar Pranowo, Ridwan Kamil, Prabowo Subianto, Sandiaga Uno, Agus Harimurti Yudhoyono dan Puan Maharani. Sumber data yang akan digunakan dari media sosial adalah Twitter. Pengguna media sosial yang ada pada Indonesia saat ini sudah mencapai 170 juta pengguna di awal tahun 2021, dengan media sosial yang cukup favorit adalah twitter dengan pembuktiannya dengan hasil dari survei bahwa Indonesia menempati sebagai posisi ke-6 dengan peringkat dunia penggunaan twitter, selain dari itu di Indonesia pengguna dengan jenjang umur 16 – 64 tahun mencapai tingkat 63,6% yang bersesuaian dengan sudah layak menjadi pemilih dalam pilpres [1].

Data dari twitter selanjutnya akan diklasifikasikan kedalam tiga kelas yakni positif, negatif, dan netral dengan menggunakan Algoritma Support Vector Machine (SVM). Penelitian ini melakukan analisis sentiment publik terkait bakal calon presiden 2024 menggunakan metode support vector machine dan didapatkan hasil bahwa akurasi tertinggi memiliki persentase 98,33% dan nilai akurasi terendah 76,67% sedangkan nilai recall tertinggi yaitu 0,90% dan nilai precision tertinggi 1,00% dimana nilai tersebut mewakili beberapa kandidat bakal calon presiden 2024.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Supervise Learning* (SVM). Metode penelitian ini merupakan jenis metode penelitian digunakan untuk membangun suatu model klasifikasi. Algoritma ini mencari hyperplane terbaik diantara beberapa kelas. penelitian ini menggunakan pendekatan One Against All (OAA) sebagai metode untuk menentukan kelas [2].

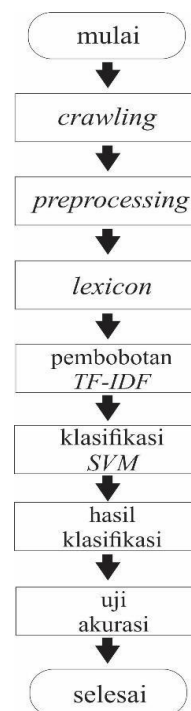
Pengambilan data pada penelitian ini lakukan dengan cara penambahan data dari twitter, dengan menggunakan API twitter dengan bantuan tools dari jupyter notebook dengan Bahasa pemrograman python. Pada tahap awal untuk mendapatkan API dilakukan dengan mendaftarkan ke Twitter developer maka akan didapatkan access token. Dengan tools jupyter notebook access token akan

diinputkan. Proses crawling menggunakan beberapa kata kunci dengan tweet berbahasa Indonesia dan disimpan dalam bentuk csv.

2.1 Tahapan Penelitian

Tahapan penelitian yang digunakan untuk menyelesaikan penelitian terdiri dari beberapa tahapan, diagram alir keseluruhan pada penelitian ini dapat dilihat pada Gambar 1.

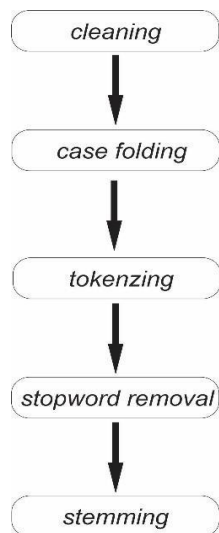
a. Gambar Tahapan Penelitian



Gambar 1 Tahapan Penelitian

Dalam tahap *preprocessing* data akan diproses kedalam 5 tahapan yakni case folding dimana semua kata dalam teks akan dirubah menjadi huruf kecil. Kemudian cleansing, pada tahap ini seluruh tanda baca koma (,), titik (.) beserta tanda baca lainnya akan dihapus. Tokenizing, pada tahapan ini seluruh teks awal yang berupa kalimat akan dipisah menggunakan koma (,) menjadi kata per kata. Stopword removal, pada tahap ini setiap kata dalam teks yang tidak memiliki arti akan dihapus. Stemming, tahapan terakhir preprocessing tahapan ini menghapus setiap imbuhan pada pada kata.

Tahap *preprocessing* penelitian ini dapat dilihat pada Gambar 2.



Gambar 2 preprocessing

2.2 Support Vector Machine (SVM)

SVM adalah metode learning machine yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. SVM merupakan cabang ilmu statistika dengan model yang akan memberikan hasil yang lebih baik dibandingkan metode lainnya [3].

Pola dari algoritma Support Vector Machine adalah hyperplane atau garis pemisah, yang mana garis tersebut memiliki peran memisahkan kelas. Persamaan perhitungan hyperplane dapat dilihat dibawah:

Hyperplane klasifikasi linear SVM dinotasikan dengan :

$$f(x)w \cdot x + b = 0 \quad (1)$$

dari persamaan diatas didapatkan pertidaksamaan kelas +1 (negatif) :

$$w \cdot x + b \leq +1 \quad (2)$$

$$\text{pertidaksamaan kelas } -1 \\ w \cdot x + b \geq -1 \quad (3)$$

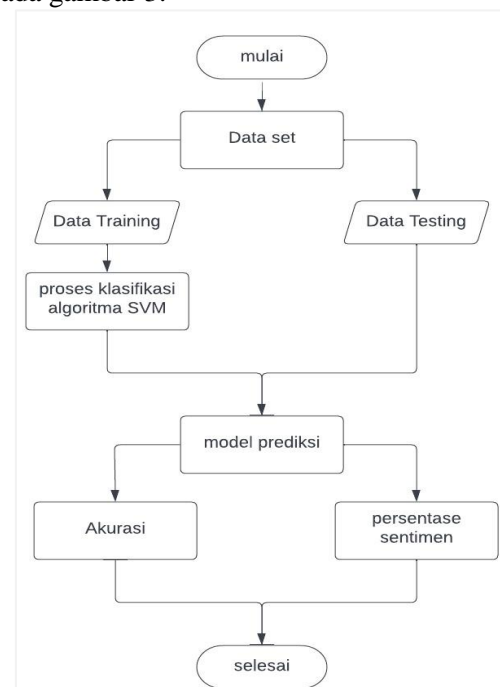
w adalah bidang normal dan b adalah posisi bidang relatif terhadap koordinat pusat. Dengan mengoptimalkan nilai jarak antara hyperplane dan titik berikutnya, margin terbesar dapat ditemukan, yaitu $1 / \|w\|$. dimana titik minimum persamaan (4) dengan mengingat syarat dari persamaan (5) tersebut.

$$\min \frac{1}{2} \|w\|^2 = \min \frac{1}{2} (w_1^2 + w_2^2) \quad (4)$$

$$y_i (w \cdot x_i + b) \geq 1, i = 1, 2, 3, \dots, N \quad (5)$$

2.3 Tahapan Klasifikasi SVM

Pada tahapan ini data dibagi menjadi 2 yakni data training dan data testing, dimana data training sebanyak 80% digunakan untuk membangun suatu model klasifikasi SVM dan data testing 20% digunakan untuk menguji model yang telah dibangun. Ketika data testing melalui proses dalam sistem dan melakukan proses klasifikasi sesuai dengan model yang telah dibangun, dan output berupa positif dan negatif. Tahapan klasifikasi SVM dapat dilihat pada gambar 3.



Gambar 3 klasifikasi SVM

2.4 Pengumpulan Data

Pengambilan data pada penelitian ini dilakukan dengan cara penambangan data dari twitter, dengan menggunakan API twitter dengan bantuan tools dari jupyter notebook dengan Bahasa pemrograman *python*. Pada tahap awal untuk mendapatkan API dilakukan dengan mendaftarkan ke Twitter *developer* maka akan didapatkan *access token*. Dengan *tools jupyter notebook access token* akan diinputkan.

Proses *crawling* menggunakan beberapa kata kunci dengan tweet berbahasa Indonesia

dan disimpan dalam bentuk csv. Adapun tokoh yang digunakan dalam penelitian ini antara lain (Ganjar Pranowo, Anies Baswedan, Ridwan Kamil, Agus Trimurthy Yudoyono, Sandiaga Uno, Prabowo Subianto, Puan Maharani).

3. HASIL DAN PEMBAHASAN

3.1 Analisis Kebutuhan

Dalam penelitian ini dilakukan analisis yang dimulai dari tahapan pengumpulan data (*crawling*) data yang diambil dari aplikasi tweeter, proses *preprocessing*, menghitung bobot perkatanya, selanjutnya pengklasifikasian sehingga mendapatkan *ouput* dalam penelitian ini.

Penelitian analisis sentiment positif, negatif dan netral komentar video tweeter diimplementasikan menggunakan *software jupyter notebook* dengan spesifikasi *Processor Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz 2.40 GHz* dan *memory 8192MB RAM* dengan *Operating System Windows 11*.

3.2 Pengumpulan Data / Crawling

proses pengambilan data berupa cuitan pengguna twitter terkait beberapa nama tokoh publik yang memiliki potensi mencalonkan diri sebagai presiden dan wakil presiden pada pemilu tahun 2024. Adapun jumlah data keseluruhan yang digunakan adalah 1325 tweet.

Tabel 1 Jumlah Tweet

| Nama | Jumlah Tweet |
|--------------------------|--------------|
| Anies Baswedan | 161 |
| Ganjar Pranowo | 175 |
| Ridwan Kamil | 212 |
| Prabowo Subianto | 254 |
| Sandiaga Uno | 248 |
| Agus Harimurti Yudhoyono | 141 |
| Puan Maharani | 134 |

3.3 Preprocessing

Untuk melanjutkan keproses *preprocessing* Langkah awal yang akan dilakukan dengan *import library*, *library* yang digunakan dalam penelitian ini dilihat pada Tabel 2.

Tabel 2. Library

| |
|--|
| <code>import pandas as pd</code> |
| <code>import matplotlib.pyplot as plt</code> |
| <code>import seaborn as sns</code> |
| <code>import collections</code> |
| <code>from sklearn import model_selection</code> |
| <code>from sklearn.model_selection import cross_val_score</code> |
| <code>from sklearn.model_selection import train_test_split</code> |
| <code>from sklearn.feature_extraction.text import TfidfVectorizer</code> |
| <code>from sklearn.metrics import accuracy_score</code> |
| <code>from sklearn.metrics import confusion_matrix</code> |
| <code>import nltk</code> |

Tujuan dilakukan *preprocessing* dalam penelitian ini untuk menyeleksi data teks agar lebih tersruktur dan menghilangkan kata yang tidak diperlukan. Adapun tahapan dalam *preprocessing* ini yaitu :

- Cleaning* data atau pembersihan data, adalah tahap pertama dalam proses analisis sentimen yang melibatkan penghapusan atau koreksi kesalahan, duplikasi, atau data yang tidak relevan dari set data yang dikumpulkan. Tahap ini penting untuk memastikan kualitas data yang akan digunakan dalam analisis sentimen.
- Case Folding* Pada tahap *case folding* pada tahap *preprocessing* data dilakukan dengan tujuan agar menyeragamkan bentuk huruf atau kata ke dalam huruf kecil serta menghilangkan tanda baca yang ada dalam kalimat ulasan. Hal tersebut dilakukan agar kata yang memiliki huruf kecil dan besar tidak terdeteksi atau diartikan memiliki perbedaan arti.
- Tokenizing* adalah proses untuk memecah kalimat yang terdapat dalam dokumen text dengan cara memisahkan kata berdasarkan kata per-spasi. Potongan kata atau kata yang sudah terpisahkan disebut sebagai token. Pemisahan kata ini akan dipisahkan dengan tanda koma (,) untuk setiap katanya.

- d) *Filtering/Stopword Removal* Filtering merupakan proses membuang kata yang tidak memiliki arti dan pengaruh terhadap analisis sentimen namun seringkali muncul dalam dokumen. Proses *filtering* disebut dengan *Stopword Removal*. Proses untuk menghilangkan kata-kata tersebut dilakukan dengan library NLTK, sebuah library yang memiliki kumpulan kata *stopword removal* dalam bahasa Indonesia.
- e) *Stemming* adalah suatu teknik dalam *preprocessing* data yang digunakan untuk menghilangkan berbagai variasi kata yang memiliki akar kata yang sama sehingga hanya menyisakan kata dasarnya. *Stemming* dapat digunakan untuk mengurangi jumlah fitur atau kata dalam suatu dataset dan memudahkan analisis data.

3.3 Pelabelan / Lexion

Data yang dihasilkan dalam proses *crawling* akan dilabeli secara menggunakan lexicon, pelabelan teks diubah kedalam bentuk *numerik* agar sesuai dengan kebutuhan sistem. Sentiment netral diberi label 2, sentiment positif diberi label 1, dan sentiment negative diberi label 0. Hasil dari pelabelan adalah data yang disimpan dalam bentuk *csv*.

Berdasarkan hasil lexicon menunjukkan bahwa kalimat “Nasdem pilih tetap Ganjar Pranowo” memiliki hasil lexicon yaitu 8 dan diberi label 1 yang berarti positif, kalimat kedua “Rindu sosok deklarasi Prabowo pilpres” memiliki hasil lexicon -1 dan diberi label 0 yang berarti negative, sedangkan kalimat ketiga “Ekonomi kreatif Sandiaga Uno nilai” memiliki hasil lexicon yaitu 0 dan diberi label 2 yang berarti netral.

3.4 Pembobotan TF-IDF

Pembobotan TF-IDF (*term Frequency-Inverse Document Frequency*) untuk melihat berapa nilai frekuensi kemunculan terhadap suatu kata. Hasil yang didapatkan dalam pembobotan TF-IDF merupakan hasil perkalian dari TF dikalikan dengan IDF.

Berdasarkan hasil *tf.idf* menunjukkan bahwa Term “Ganjar” memiliki nilai *idf* yaitu 0, Term “Pranowo” memiliki nilai *idf* yaitu

0.031, Term “Tetap” memiliki nilai *idf* yaitu 0.602, Term “Pilpres” memiliki nilai *df* yaitu 0.031, Term “Pilih” memiliki nilai *df* yaitu 0.602.

Keterangan : Term : kalimat

TF-IDF : hasil pembobotan

Hasil pembobotan TF-IDF merupakan data numerik yang dimana nantinya data tersebut digunakan dalam proses perhitungan klasifikasi SVM.

3.5 Pembagian Data

Pada tahap proses klasifikasi SVM , penelitian ini menggunakan SVM kernel linear dan dibangun sebuah *machine learning* dengan menggunakan perbandingan data 80% data *training* dan 20% data *testing* dari sumber data *tweet* dengan jumlah data yang digunakan seperti berikut:

Tabel 3 Pembagian Data

| No | Nama Kandidat | Data Training | Data Testing | Total |
|----|--------------------------|---------------|--------------|-------|
| 1 | Anies Baswedan | 128 | 33 | 161 |
| 2 | Ganjar Pranowo | 140 | 35 | 175 |
| 3 | Ridwan Kamil | 169 | 43 | 212 |
| 4 | Prabowo Subianto | 203 | 51 | 254 |
| 5 | Sandiaga Uno | 198 | 50 | 248 |
| 6 | Agus Harimurti Yudhoyono | 112 | 29 | 141 |
| 7 | Puan Maharani | 107 | 27 | 134 |

3.6 Pengujian

Terdapat tujuh kandidat yang dinilai berdasarkan tiga metrik performa tersebut. Kandidat-kandidat tersebut adalah Anies Baswedan, Ganjar Pranowo, Ridwan Kamil, Prabowo Subianto, Sandiaga Uno, Agus Harimurti Yudhoyono, dan Puan Maharani. Masing-masing kandidat memiliki nilai *accuracy*, *recall*, dan *precision* yang berbeda-beda. Terlihat bahwa kandidat Ridwan Kamil dan Ganjar Pranowo memiliki nilai *accuracy* yang paling tinggi, sedangkan kandidat Anies Baswedan memiliki nilai *accuracy* yang paling rendah.

Tabel 4 Hasil Pengujian Akurasi

| No | Nama Kandidat | Accuracy | Recall | Precision |
|----|--------------------------|----------|--------|-----------|
| 1 | Anies Baswedan | 63.64 % | 0.25 | 0.60 |
| 2 | Ganjar Pranowo | 85.71 % | 0.00 | 0.00 |
| 3 | Ridwan Kamil | 90.7 % | 0.43 | 1.00 |
| 4 | Prabowo Subianto | 76.47 % | 0.41 | 0.88 |
| 5 | Sandiaga Uno | 70.0 % | 0.56 | 0.90 |
| 6 | Agus Harimurti Yudhoyono | 75.86 % | 0.79 | 0.92 |
| 7 | Puan Maharani | 70.37 % | 0.44 | 1.00 |

Tabel di atas menunjukkan hasil pengujian klasifikasi SVM pada 7 kandidat dengan parameter akurasi, *recall*, dan presisi. Akurasi mengukur seberapa tepatnya model dalam mengklasifikasikan keseluruhan data. *Recall* mengukur seberapa banyak kelas positif yang berhasil diidentifikasi oleh model dari keseluruhan kelas positif yang sebenarnya. Sedangkan presisi mengukur seberapa banyak kelas positif yang diidentifikasi oleh model yang benar dari keseluruhan hasil identifikasi kelas positif. Dari segi akurasi, kandidat Ridwan Kamil memiliki nilai tertinggi yaitu 90.7%, yang artinya model berhasil mengklasifikasikan 90.7% data dengan benar. Sedangkan dari segi *recall*, kandidat Agus Harimurti Yudhoyono juga memiliki nilai tertinggi yaitu 0.79, yang artinya model berhasil mengidentifikasi 79% kelas positif dari keseluruhan kelas positif yang sebenarnya.

3.6 Hasil Sentimen

Tabel 5 Hasil Sentimen

| No | Bakal Calon | Positive | Netral | Negative |
|----|--------------------------|----------|--------|----------|
| 1 | Anies Baswedan | 84.85% | 0.00% | 15.15% |
| 2 | Ganjar Pranowo | 100.00% | 0.00% | 0.00% |
| 3 | Ridwan Kamil | 88.37% | 4.65% | 6.98% |
| 4 | Prabowo Subianto | 84.31% | 0.00% | 15.69% |
| 5 | Sandiaga Uno | 76.00% | 4.00% | 20.00% |
| 6 | Agus Harimurti Yudhoyono | 55.17% | 3.45% | 41.38% |
| 7 | Puan Maharani | 85.19% | 6.67% | 14.81% |

Dari tabel tersebut dapat disimpulkan bahwa Ganjar Pranowo adalah tokoh yang paling banyak mendapatkan sentimen positif dari responden, sedangkan Puan Maharani adalah tokoh yang paling banyak mendapatkan sentimen negatif. Namun demikian, untuk menginterpretasikan hasil analisis sentimen dengan lebih tepat, perlu diketahui juga mengenai metode yang digunakan dalam pengumpulan data dan pengolahan data.

4. KESIMPULAN

Terdapat perbedaan performa antara setiap kandidat dalam pengujian atau pengukuran yang dilakukan dengan menggunakan metrik-metrik seperti *accuracy*, *recall*, dan *precision*. Kandidat Ridwan Kamil memiliki nilai *accuracy* tertinggi dengan persentase 90.7%, sementara kandidat Anies Baswedan memiliki nilai *accuracy* terendah sebesar 63.64 %. Dari segi *precision*, kandidat Ridwan Kamil dan, Puan Maharani masing-masing memiliki nilai *precision* tertinggi dengan persentase 1.00. Dan dari segi *Recall* Agus Harimurti Yudhoyono memiliki nilai *Recall* tertinggi dengan persentase 0.79. Sehingga, kesimpulan yang dapat diambil adalah terdapat kandidat yang memiliki performa tinggi secara keseluruhan (dilihat dari *accuracy*), namun mungkin kurang baik dalam mengidentifikasi data positif (dilihat dari *recall*) atau memprediksi data positif secara akurat (dilihat dari *precision*). Oleh karena itu, perlu dilakukan evaluasi yang holistik dan mendalam dalam memilih kandidat yang terbaik dalam situasi dan kondisi tertentu.

5. SARAN

Diharapkan kedepannya proses pelabelan bisa dilakukan secara otomatis. Sebaiknya juga dilakukan dengan menggunakan metode lain untuk melihat metode mana yang lebih akurat.

UCAPAN TERIMA KASIH

Terimakasih kepada Dekan fakultas teknik Dr. Ir. Hafsah Nirwana, M.T., Ketua Program Studi Informatika Muh. Fahmi Rustan, S.kom., MT. Rekan – rekan kerja di Lab Sistem Cerdas yang telah sabar untuk

memberikan masukan hingga penelitian ini dapat terselesaikan.

DAFTAR PUSTAKA

- [1] Muhammad Raihan Fals sya' bani' Ultach Enri, Tesa Nur Padilah (2022). Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes.
 - [2] Dhina Nur Fitriana and Yuliant Sibaroni (2020). Sentiment Analysis On Kai Twitter Post Using Multiclass Support Vector Machine (SVM)
 - [3] Darwis, D., Pratiwi, E.S. and Pasaribu, A.F.O. (2020) 'Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia', *Edutic - Scientific Journal of Informatics Education*, 7(1), pp. 1–11. Available at: <https://doi.org/10.21107/edutic.v7i1.8779>.
-

