

## Analisis Sentimen Terhadap Film *Dirty Vote* Pada Platform Twitter (X) Dengan Metode *Naïve Bayes* Dan *Cosine Smiliarity*

Mega Ihsani Rahmat<sup>1</sup>, Asriyanik<sup>2</sup>, Agung Pambudi<sup>3</sup>  
<sup>1,2,3</sup> Universitas Muhammadiyah Sukabumi, Sukabumi  
Email : megaihsanirahmat09@gmail.com

### Abstrak

Penelitian bertujuan untuk menganalisis sentimen pengguna media sosial terhadap film "*Dirty Vote*" menggunakan metode text mining dan algoritma *Naïve Bayes*. Data diambil melalui teknik *crawling* dari Twitter (X) untuk mendapatkan tweet yang berkaitan dengan film tersebut. Setelah data dikumpulkan, dilakukan pelabelan secara manual untuk menentukan apakah sentimen yang terkandung dalam tweet tersebut bersifat positif atau negatif. Proses *pre-processing* dilakukan untuk membersihkan data dan menyiapkannya untuk analisis lebih lanjut, termasuk tokenisasi dan stemming. Pembobotan kata menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) diterapkan untuk menentukan pentingnya kata-kata dalam konteks dokumen. Algoritma *Naïve Bayes* kemudian digunakan untuk memodelkan dan mengklasifikasikan sentimen. Evaluasi model dilakukan dengan *Confusion Matrix* dan *K-fold Cross Validation* untuk memastikan akurasi hasil yang diperoleh. Hasil penelitian menunjukkan bahwa model yang dibangun memiliki akurasi yang cukup tinggi dalam mengklasifikasikan sentimen, memberikan pengetahuan yang berguna bagi pembuat film dan pemasar dalam memahami persepsi publik terhadap film "*Dirty Vote*".

**Kata Kunci:** *Analisis Sentimen, Film Dirty Vote, Twitter (X), Naive Bayes, Cosine Similarity.*

### PENDAHULUAN

Film memainkan peran penting dalam budaya dan hiburan masyarakat. *Review* dan opini mengenai film kini dapat dengan mudah ditemukan di berbagai platform media sosial dan situs *review* berkat kemajuan teknologi dan internet. Analisis sentimen terhadap ulasan film menjadi alat yang penting untuk memahami pandangan penonton, mengukur kepuasan, dan memprediksi kesuksesan sebuah film. Salah satu film yang menarik untuk dikaji adalah "*Dirty Vote*", telah mendapatkan perhatian luas sehingga mempengaruhi beberapa opini khalayak banyak terutama pengguna twitter.

Twitter (X) adalah platform mikroblog yang memungkinkan pengguna berkomunikasi melalui pesan singkat (Fikri et al., 2020). Fitur penting dari Twitter (X) adalah bagian trendingnya, di mana topik yang mendapat perhatian besar dari pengguna cenderung menjadi populer dan

menghasilkan diskusi berkelanjutan. Antarmuka platform yang ramah pengguna telah berkontribusi terhadap penerapannya secara luas di kalangan masyarakat Indonesia, yang memanfaatkan Twitter (X) sebagai sumber informasi utama (Styawati et al., 2021). Saat ini, topik yang cukup menarik perhatian adalah film kontroversial "*Dirty Vote*".

*Dirty vote* merupakan film yang berisi tentang kecurangan yang terjadi pada saat pemilu. Film ini menampilkan tiga pakar hukum yang menjelaskan bagaimana instrumen-instrumen kekuasaan dipakai untuk mencurangi pemilu yang sudah pasti akan merusak tatanan demokrasi negara kita. Film ini mengungkap bahwa adanya kecurangan yang dilakukan para politisi untuk kepentingan pribadi mereka. Beberapa tokoh politik juga disebut terkait pembahasan tentang penunjukan pejabat daerah yang dianggap sebagai praktik politik balas budi dan berusaha

menciptakan loyalitas terhadap mereka yang berkuasa, pemekaran provinsi di Papua, penyalangan penyaluran bantuan sosial di beberapa daerah oleh pejabat publik yang diduga bantuan sosial dimanfaatkan untuk meningkatkan suara di pemilu bahkan meningkat tajam jika dibandingkan dengan pada saat masa pandemi, kontroversi penggunaan fasilitas negara untuk kampanye, bawaslu yang dianggap gagal dalam mengawal adanya pemilu, dan juga ketidaknetralan presiden. Ketiga narasumber di film ini juga mengungkap timeline apa yang sebenarnya terjadi dari awal sampai detik-detik akhir di putusan Mahkamah Konstitusi dalam meloloskan salah satu wakil presiden yang menyalahi prosedur. Ada beberapa aspek yang membuat film ini menarik untuk dilihat atau setidaknya dipahami oleh orang yang akan melakukan pemilu di masa yang akan datang untuk mempertimbangkan kembali hal-hal yang masih bisa dipertimbangkan sebelum akhirnya memilih presiden di masa yang akan datang.

Analisis sentimen adalah metode untuk mengidentifikasi dan mengekstraksi opini atau sentimen dari teks. Dalam konteks film "*Dirty Vote*", analisis sentimen dapat digunakan untuk mengklasifikasikan komentar atau ulasan penonton menjadi klasifikasi perasaan positif, negatif, atau nonpartisan (netral). (Giovani et al., 2020)

Metode *Naive Bayes* adalah salah satu algoritma pembelajaran mesin yang populer untuk klasifikasi teks, termasuk analisis sentimen (Styawati et al., 2021). Algoritma ini didasarkan pada Teorema Bayes dengan asumsi bahwa fitur-fitur (kata-kata) dalam sebuah teks saling independen. Keuntungan utama dari *Naive Bayes* adalah kesederhanaan dan efisiensinya dalam menangani dataset besar, meskipun asumsi independensinya sering kali tidak realistis (Darwis et al., 2021).

*Cosine Similarity* adalah teknik yang digunakan untuk mengukur kemiripan antara dua vektor dalam ruang sorotan. Dalam analisis teks, vektor ini biasanya

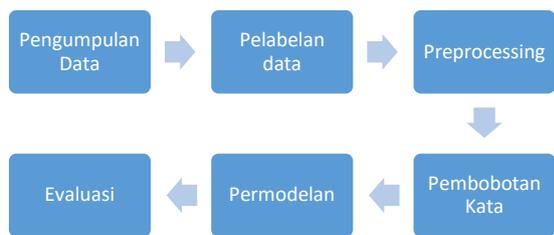
mewakili dokumen yang diubah menjadi representasi numerik menggunakan teknik seperti *bag-of-words* atau TF-IDF (AZMI et al., 2020). *Cosine Similarity* mengukur sudut antara dua vektor, memberikan nilai antara -1 dan 1, dimana nilai 1 berarti vektor tersebut identik dan nilai 0 berarti tidak ada kesamaan. Metode ini berguna untuk mengukur kemiripan antara ulasan-ulasan yang ada dan membantu dalam *clustering* atau pengelompokan ulasan berdasarkan kesamaan (M Yunus, 2020).

Dalam pembahasan film "*Dirty Vote*", analisis sentimen menggunakan metode *Naive Bayes* dan *Cosine Similarity* dapat memberikan wawasan yang berharga tentang bagaimana film tersebut diterima oleh penonton. Dengan mengumpulkan ulasan dari berbagai sumber, kita dapat melatih model *Naive Bayes* untuk mengklasifikasikan sentimen ulasan sebagai positif dan negatif. Selain itu, dengan menggunakan *Cosine Similarity* dapat mengelompokkan ulasan-ulasan yang serupa, mengidentifikasi tema umum, dan mendapatkan gambaran yang lebih baik tentang sentimen masyarakat.

*Naive Bayes* memiliki tingkat presisi yang lebih tinggi, sehingga pengklasifikasi *Naive Bayes Classifier* yaitu proses karakterisasi teks dengan mempertimbangkan kemungkinan kehadiran kata-kata. Berdasarkan pertimbangan tersebut, penulis tertarik untuk melakukan penelitian mengenai sentimen masyarakat terhadap film *Dirty Vote* khususnya di Twitter (X), untuk mengetahui apakah film ini dipandang positif atau negatif. Sehingga eksplorasi ini diberi nama "Analisis Sentimen Terhadap Film *Dirty Vote* Pada Platform Twitter Dengan Metode *Naive Bayes* dan *Cosine Similarity*".

## METODE

Dibawah ini adalah tahapan dalam penelitian:



Gambar 1. Tahapan Penelitian

- **Pengumpulan Data**  
Data yang digunakan dalam penelitian ini bersumber dari Twitter (X), dimana pengumpulan data dilakukan *crawling* dengan menggunakan bahasa pemrograman Python untuk mengakses Twitter API dan mengumpulkan informasi terkait film dirty vote.
- **Pelabelan Data**  
Selama fase ini, data pelatihan dianotasi dengan setiap teks dokumen yang ditetapkan sebagai sentimen positif atau negatif. Tahapan dilakukan secara manual.
- **Pre-Procesing**  
Fase selanjutnya melibatkan *pre-processing*, sebuah langkah penting yang bertujuan menghilangkan atribut yang tidak perlu dalam persiapan untuk proses klasifikasi. Fase ini mencakup serangkaian prosedur, termasuk *cleaning*, *case folding*, *stopword*, identifikasi kata *slangword*, *tokenizing*, dan *stemming*, untuk menangani data mentah.
  - a) **Cleaning**  
Langkah pertama dalam pra-pemrosesan melibatkan pemeriksaan data untuk mencari potensi masalah seperti entri duplikat, nilai yang hilang, data yang salah, atau informasi yang tidak relevan. Untuk menyederhanakan proses analisis sentimen, atribut asing seperti sebutan, hashtag, dan tautan akan dikecualikan.

Tabel 1. *Cleaning*

Sebelum	Sesudah
@belswen “yg disampaikan di Dirty Vote bernada fitnah, narasi kebencian, yg asumtif dan tidak ilmiah”	yg disampaikan di Dirty Vote bernada fitnah narasi kebencian yg asumtif dan tidak ilmiah

- b) **Case Folding**  
Tujuan dari fase ini adalah mengubah huruf besar yang ada dalam teks komentar atau tweet menjadi huruf kecil.

Tabel 2. *Case Folding*

Sebelum	Sesudah
yg disampaikan di Dirty Vote bernada fitnah narasi kebencian yg asumtif dan tidak ilmiah	yg disampaikan di dirty vote bernada fitnah narasi kebencian yg asumtif dan tidak ilmiah

- c) **Tokenizing**  
Proses tokenisasi melibatkan segmentasi teks menjadi unit-unit berbeda dengan memisahkan kata-kata yang dibatasi oleh spasi. Tugas ini dilakukan dengan menggunakan *Library Natural Language Toolkit* dalam bahasa pemrograman Python.

Tabel 3. *Tokenizing*

Sebelum	Sesudah
yg disampaikan di dirty vote bernada fitnah narasi kebencian yg asumtif dan tidak ilmiah	'yg' 'disampaikan' 'di' 'dirty' 'vote' 'bernada' 'fitnah' 'narasi' 'kebencian' 'yg' 'asumtif' 'dan' 'tidak' 'ilmiah'

- d) **Stopword**  
Penghapusan *stopword* adalah teknik yang digunakan untuk menghilangkan kata-kata tidak penting, juga dikenal sebagai kata-kata berhenti, dari kumpulan data teks. Kata-kata ini, yang biasanya mencakup kata penghubung umum, tidak berdampak signifikan terhadap keakuratan proses klasifikasi. Dalam konteks penghapusan *stopword*, kata-kata tersebut diganti dengan spasi, sedangkan kata-kata penting

dipertahankan dalam daftar kata tersendiri.

**Tabel 4. Stopword**

Sebelum	Sesudah
'yg' 'disampaikan' 'di' 'dirty' 'vote' 'bernada' 'fitnah' 'narasi' 'kebencian' 'yg' 'asumtif' 'dan' 'tidak' 'ilmiah'	'disampaikan' 'dirty' 'vote' 'bernada' 'fitnah' 'narasi' 'kebencian' 'asumtif' 'ilmiah'

- e) Slangword  
Proses ini berupaya untuk menggantikan istilah-istilah sehari-hari atau informal yang sering digunakan dalam interaksi sehari-hari dengan istilah-istilah formal.

**Tabel 5. Slangword**

Sebelum	Sesudah
Lg	Lagi
Yg	Yang
Dlm	Dalam

- f) Stemming  
Tujuan dari prosedur ini adalah untuk menghilangkan imbuhan linguistik dan mengembalikan kata tersebut ke bentuk akarnya. Proses stemming bahasa Indonesia melibatkan pemanfaatan *Library Sastrawi* bahasa pemrograman Python untuk menghilangkan prefiks dan sufiks.

**Tabel 6. Stemming**

Sebelum	Sesudah
disampaikan dirty vote bernada fitnah narasi kebencian asumtif ilmiah	sampai dirty vote nada fitnah narasi benci asumtif ilmiah

- **Pembobotan Kata**  
Pada tahap ini, tujuannya adalah untuk mengurangi data yang tidak relevan dan sekaligus meningkatkan data yang relevan guna mencapai klasifikasi yang lebih akurat. Untuk mencapainya, diterapkan teknik Term Frekuensi-Invers Dokumen Frekuensi (TF-IDF), yang menghitung frekuensi kemunculan setiap istilah dalam dokumen dan

seluruh koleksi dokumen. TF-IDF merupakan suatu proses untuk menilai tingkat kepentingan suatu kata (term) dalam dokumen serta korpus secara keseluruhan. Langkah ini bertujuan untuk menyoroti bobot atau relevansi kata dalam konteks dokumen tertentu dibandingkan dengan keseluruhan dokumen dalam korpus.

$$IDF(t) = \log \frac{N}{df(t)}$$

Keterangan:

TF = Teks frekuensi

IDF = Teks frekuensi dalam dokumen

N = Jumlah dokumen

- **Pemodelan Algoritma**

Menurut (Murni et al., 2023) Metode Naïve Bayes Classifier untuk analisis sentimen berdasarkan pada teorema bayes seperti berikut:

$$P(C|X) = \frac{P(X|C) * P(C)}{P(X)}$$

Keterangan:

P(C|X) = Probabilitas posterior dari kelas C yang diberikan masukan X

P(X|C) = Probabilitas bersyarat dari masukan X yang diberikan kelas C

P(C) = Probabilitas prior dari kelas C

P(X) = Probabilitas prior dari masukan X

- **Evaluasi**

a) Confusion Matrix

Confusion Matrix merupakan table yang menunjukkan empat kombinasi antara hasil prediksi dan nilai aktual dalam proses klasifikasi. Model evaluasi kinerja dilakukan dengan menggunakan dua metrik utama, yaitu F1-Score yang menggabungkan presisi dan recall secara seimbang, serta akurasi yang mengukur klasifikasi secara keseluruhan. F1-Score memberikan Gambaran menyeluruh tentang model efektivitas, sementara akurasi menunjukkan persentase prediksi yang benar dari total kasus yang ada (Murni et al., 2023)

Tabel 7. Confusion Matrix

Confusion Matrix	Predicted Class	
Actual Class	Positive	Negative
Positive	TP	FN
Negative	FP	TN

- b) **K-Fold Cross Validation**  
K-Fold Cross Validation adalah metode evaluasi yang digunakan untuk mengukur efektivitas model pembelajaran mesin. Teknik ini membagi data menjadi K subset yang memiliki ukuran yang sama secara acak. Validasi dilakukan sebanyak K kali, dimana setiap subset akan berfungsi sebagai data uji sekali, sementara subset lainnya digunakan sebagai data latih. Pada setiap iterasi, model dibor dengan data latih dan kinerjanya dievaluasi.

## HASIL DAN PEMBAHASAN

- **Pengumpulan Data**  
Dataset yang digunakan dalam penelitian ini terdiri dari tweet dan komentar berbahasa Indonesia yang diperoleh dari X (Twitter) dengan kata kunci “Film Dirty Vote”. Data dikumpulkan melalui proses crawling menggunakan bahasa pemrograman Python di Google Collab dengan bantuan Twitter auth token. Proses pencarian dilakukan antara tanggal 11 hingga 18 Februari 2024, yakni dari tanggal perilis film hingga tujuh hari setelahnya. Data yang terkumpul disimpan dalam format CSV, kemudian dilakukan pembersihan dengan menghapus duplikat serta data yang tidak relevan dan tidak bermakna secara manual. Hasilnya, sebanyak 1000 data bersih.
- **Pelabelan Data**  
Pada tahapan ini, setiap dokumen dalam kumpulan data diberi klasifikasi. Klasifikasi yang digunakan mencakup dua kategori berbeda: komentar positif dan komentar negatif. Proses pelabelan data yang berdasarkan dua kategori ini menghasilkan masing-masing data per-kategori sebagai berikut.

Tabel 8. Jumlah Pelabelan Data

No	Label	Jumlah
1	Positif	537
2	Negatif	263
<b>Jumlah Keseluruhan</b>		800

- **Pre-Processing**  
Proses Pre-Processing dimulai dengan tahap pembersihan data (cleaning), kemudian dilanjutkan dengan casefolding, tokenizing, penghapusan stopwords, penanganan slangword, dan diakhiri dengan stemming. Semua tahap bantuan ini dilakukan menggunakan bahasa pemrograman Python dengan beberapa perpustakaan, seperti yang ditampilkan berikut.

```
import re, string
import pandas as pd
import numpy as np
import nltk
import csv
import time
from datetime import timedelta
import json
from nltk.tokenize import word_tokenize
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.probability import FreqDist
from google.colab import drive
drive.mount('/content/gdrive')
```

Gambar 2. Library dan Source Code Pre-Processing

- **Pembobotan Kata**  
Selanjutnya, menentukan signifikansi sebuah kata dalam sebuah kalimat dengan menerapkan algoritma TF-IDF menggunakan library `sklearn.feature_extraction.text`.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report
from sklearn.model_selection import train_test_split as split
from sklearn.model_selection import KFold
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score
from tabulate import tabulate
import numpy as np
import pickle
```

Gambar 3. Library TF-IDF

- **Pemodelan Algoritma**  
Setelah data diproses sesuai kebutuhan model, data dibagi menjadi dua bagian: data pelatihan dan data pengujian. Data training digunakan untuk melatih model agar dapat mempelajari pola dan hubungan dalam data, sementara data testing digunakan untuk memancarkan kinerja model dengan menguji kemampuannya dalam memprediksi data yang belum pernah dilihat sebelumnya. Pada tahap klasifikasi menggunakan Naive Bayes

Multinomial, ada beberapa library yang diperlukan diantaranya sebagai berikut.

```

for train_index, test_index in kf.split(data):
    X = data['tweet'][train_index]
    Y = data['Sentimen'][train_index]

    x_train, x_test, y_train, y_test = split(X, Y, test_size=0.2, random_state=42)

    vectorizer = TfidfVectorizer()
    train_vector = vectorizer.fit_transform(x_train)
    test_vector = vectorizer.transform(x_test)

    clf = MultinomialNB()
    clf_train = clf.fit(train_vector, y_train)

    save_model = open('/content/gdrive/MyDrive/Giani/model_NBC.pickle', 'wb')
    pickle.dump(clf_train, save_model)
    save_model.close()

    clf_model = open('/content/gdrive/MyDrive/Giani/model_NBC.pickle', 'rb')
    clf = pickle.load(clf_model)

    y_pred = clf.predict(test_vector)

    print('Precision : ', precision_score(y_test, y_pred, average = 'micro'))
    print('Recall : ', recall_score(y_test, y_pred, average = 'micro'))
    print('F1-Score : ', f1_score(y_test, y_pred, average = 'micro'))
    
```

Gambar 4. Source Code Naïve Bayes Classifier

- **Evaluasi**

a) Confusion Matrix

Hasil dari proses klasifikasi menggunakan Naïve Bayes dievaluasi dengan perhitungan presisi, recall, dan F1-Score. Tujuan dari evaluasi ini adalah untuk mengukur akurasi kinerja algoritma Naïve Bayes Multinomial pada dataset.

Tabel 9. Hasil Confusion Matrix

	Precision	Recall	F1-Score
Pro	0.66	0.99	0.80
Kontra	0.88	0.12	0.21

Dapat dilihat seperti pada tabel diatas, untuk label positif sistem dapat memperoleh akurasi antara informasi yang tersedia sebesar 66% sedangkan untuk tabel negatif sebesar 88%. Tingkat keberhasilan sistem dalam memulihkan kembali informasi adalah sebesar 99% untuk label positif dan 12% untuk label negative.

b) K-Fold Cross Validation

Selanjutnya diterapkan teknik K-fold Cross Validation untuk memastikan hasil pengujian dan evaluasi kinerja algoritma mencapai hasil yang optimal. Pengujian menggunakan K-Fold Cross Validation menunjukkan sistem memiliki rata-rata tingkat akurasi pada dataset sebagai berikut.

Tabel 10. Hasil K-Fold Cross Validation

	Akurasi	Recall
Pro	67%	99%
Kontra	83%	10%

**SIMPULAN (PENUTUP)**

Penerapan metode Naive Bayes dan Cosine Similarity dalam analisis sentiment terhadap Film Dirty Vote dalam platform media sosial Twitter (X) maka dapat disimpulkan sebagai berikut :

1. Tingkat sentimen atau tanggapan pengguna di media sosial X terhadap Film Dirty Vote cenderung negatif. Dapat dilihat dari klasifikasi naive bayes classifier yang menunjukkan presentase sebesar 66%, dan presentase akurasi yang ditunjukkan oleh confusion matrik untuk pengujian yaitu sebesar adalah sebesar 68% untuk label pro dan 88% untuk label kontra.
2. Selanjutnya, hasil klasifikasi menggunakan Cosine Similarity menunjukkan bahwa kategori yang tertinggi adalah partisipasi dengan angka presentase sebesar 54,1%, lalu disusul oleh media dengan 21,9% dan Kekuasaan sebesar 17,2% serta yang terakhir adalah produksi sebesar 6,8%.
3. Penerapan K-Fold Cross Validaton menghasilkan nilai akurasi dengan nilai presentase sebesar 67% untuk komentar pro, dan akurasi dengan nilai presentase sebesar 83% untuk komentar kontra.

**UCAPAN TERIMA KASIH**

Ungkapan rasa terimakasih ingin penulis sampaikan kepada:

1. Tuhan yang Maha Esa, Allah SWT yang telah memberi kemudahan serta keridhoan-Nya selama proses penyusunan skripsi ini
2. Orang tua dan keluarga yang telah banyak memberikan kasih sayang yang tulus serta dukungan penuh baik berupa moral maupun materi. Terima kasih telah menjadi sumber motivasi penulis dalam proses penyelesaian kuliah.
3. Ibu Asriyanik, M.T. Selaku Dosen Pembimbing yang telah membimbing

dengan sangat baik dan memberikan berbagai saran dan masukan selama dalam proses pelaksanaan penelitian

4. Teman-teman baik yang telah membantu dan memberikan dorongan sehingga penulis dapat menyelesaikan laporan ini dengan baik
5. Yang terakhir, Raka Ramdani yang telah kebersamai penulis pada hari-hari yang tidak mudah selama proses penyelesaian kuliah ini dan senantiasa sabar menghadapi saya. Terima kasih telah menjadi bagian dari perjalanan saya hingga saat ini.

#### DAFTAR PUSTAKA

- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711.
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Comparison of Naïve Bayes and Support Vector Machine Methods in Twitter Sentiment Analysis. *Smatika Jurnal*, 10(02), 71–76.
- Asri, R. (2020). Membaca Film Sebagai Sebuah Teks: Analisis Isi Film “Nanti Kita Cerita Tentang Hari Ini (NKCTHI).” *Jurnal Al Azhar Indonesia Seri Ilmu Sosial*, 1(2), 74. <https://doi.org/10.36722/jaiss.v1i2.462>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155. <https://doi.org/10.30591/jpit.v6i3.2870>
- AZMI, F., Saleh, A., & Dharshinni, N. P. (2020). Face Identification on Login Security Using Algorithm Combination of Viola-Jones and Cosine Similarity. *Journal of Informatics and Telecommunication Engineering*, 4(1), 203–211. <https://doi.org/10.31289/jite.v4i1.3885>
- Telaumbanua, K., & Nababan, L. (2022). Implementasi Metode Cosine Similarity Dalam Mendeteksi Kemiripan Dan Perbedaan Gambar Hasil Scan Berbasis Android. *Informatics Engineering and Electronic Data (IEED)*, 1(1), 27–36. <https://doi.org/10.59840/ieed.v1i1.176>
- M Yunus, B. (2020). Similarity Detection for Hadith of Fiqh of Women using Cosine Similarity and Boyer Moore Method. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(1), 65–73. <https://doi.org/10.30534/ijatcse/2020/11912020>