



Analisis Sentimen Terhadap Isu Kecurangan Pemilu 2024 Pada Platfom Twitter (X) Dengan Metode *Naive Bayes Multinomial* Dan *Cosine Similiarity*

Muhammad Giani Putra^{1*}, Iwan Rizal Setiawan², Didik Indrayana³

^{1*,2,3}Program Studi Teknik Informatika, Universitas Muhammadiyah Sukabumi, Indonesia

*Email : giani_putra29@ummi.ac.id

Abstract: *In an increasingly complex digital era, sentiment analysis has become a vital instrument in understanding the nuances of public opinion. This technique, which utilizes artificial intelligence and Machine Learning, allows us to extract knowledge about people's attitudes, emotions and perceptions towards various issues. This research examines public sentiment regarding the issue of fraud in the 2024 Election on the social media platform Twitter using a text mining-based sentiment analysis approach. Data was obtained through a crawling process using the Python programming language. The research methodology includes a series of stages, starting from data cleaning to improve quality, continuing with word weighting using the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm, and ending with modeling using the Naive Bayes Classifier algorithm. Model evaluation was carried out systematically by applying the Naive Bayes, Confusion Matrix and K-Fold Cross Validation methods to measure the level of accuracy and effectiveness of the model developed. This research aims to produce in-depth knowledge regarding the trends and dynamics of public sentiment regarding the issue of fraud in the 2024 Election in the realm of social media, especially Twitter (X). Based on the research results, it shows a percentage of 67.7%, and the percentage of accuracy shown by the confusion matrix for testing is 68% for the pro label and 80% for the contra label.*

Keywords: *Naive Bayes, Cosine Similarity, Twitter (X), Sentiment Analysis*

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat serta akses internet yang semakin mudah dan terjangkau telah mendorong peningkatan penggunaan media sosial secara global setiap tahunnya. Media sosial menyajikan kemampuan bagi penggunanya untuk mengekspresikan diri, terlibat dalam percakapan, bertukar informasi, atau bahkan melakukan komunikasi secara langsung (real-time) dengan pengguna lainnya (Febriyani & Februariyanti, 2023). Melalui fitur-fitur yang disediakan, seperti membuat postingan, mengunggah foto/video, berkomentar, membagikan tautan, dan lain sebagainya sehingga pengguna dapat berinteraksi, berbagi konten, dan terhubung dengan secara lebih personal maupun publik. Sifat media sosial yang interaktif dan terhubung ini memungkinkan terciptanya komunitas virtual dengan tingkat keterlibatan dan percakapan yang tinggi di dalamnya. Menurut laporan *We Are Social*, jumlah pengguna twitter (X) di Indonesia mencapai 14,75 juta dan menduduki posisi ke-enam dibawah India pada bulan april 2023 (Murni et al., 2023)

Seiring dengan meningkatnya penggunaan media sosial, isu-isu politik seperti kecurangan pemilu menjadi topik yang sering dibicarakan oleh masyarakat. Isu kecurangan pemilu 2024, khususnya, telah memicu berbagai reaksi dan sentimen dari masyarakat yang disampaikan melalui media sosial. Sentimen publik ini penting untuk dipahami oleh berbagai pihak, termasuk pemerintah, penyelenggara pemilu, serta masyarakat itu sendiri, guna mengambil langkah-langkah yang tepat dalam menghadapi isu ini.

Analisis sentimen terhadap isu kecurangan pemilu pasca-penyelenggaraan tetap penting karena beberapa alasan. Pertama, persepsi masyarakat tentang integritas pemilu memiliki dampak jangka panjang terhadap kepercayaan publik pada sistem demokrasi dan lembaga-lembaga politiknya. Kedua, pemahaman tentang sentimen publik pasca pemilu dapat memberikan wawasan berharga bagi pembuat kebijakan dan penyelenggara pemilu untuk meningkatkan proses elektoral di masa depan.

Menurut (Florensus Sianipar et al., 2023) analisis sentimen merupakan suatu metode dalam pemrosesan bahasa alami yang bertujuan untuk mendeteksi dan mengekstraksi opini, sentimen, atau emosi seseorang terhadap suatu topik, produk, layanan, atau objek tertentu. Proses ini melibatkan pengolahan data teks untuk mengidentifikasi apakah opini yang diekspresikan cenderung positif, negatif, atau netral.

Analisis sentimen dapat dilakukan pada tingkatan yang berbeda-beda, tergantung pada tingkat detail yang diinginkan. Dua tingkatan umum yang sering digunakan adalah analisis sentimen tingkat dokumen dan analisis sentimen tingkat kalimat. Selain itu analisis sentimen tingkat dokumen (*coarse-grained sentiment analysis*) memperlakukan keseluruhan isi dokumen sebagai satu kesatuan dan mengasumsikan bahwa seluruh dokumen tersebut mengekspresikan sentimen yang bersifat positif atau negatif secara keseluruhan (Sari et al., 2020). Sementara itu, analisis sentimen tingkat kalimat (*fine-grained sentiment analysis*) adalah pendekatan yang lebih terperinci, di mana setiap kalimat dalam dokumen dianalisis secara individual untuk menentukan apakah kalimat tersebut mengandung sentimen positif, negatif, atau netral. Pendekatan ini memberikan granularitas yang lebih tinggi dalam mengidentifikasi sentimen, karena memungkinkan untuk menangkap variasi sentimen dalam satu dokumen yang sama (Ipmawati et al., 2024).

2. METODOLOGI PENELITIAN

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari Twitter (X), pengambilan data dilakukan dengan melakukan crawling (pengambilan data) menggunakan bahasa pemrograman Python dari Twitter API (Antarmuka Pemrograman Aplikasi) terkait isu kecurangan Pemilu 2024

2.2. Pelabelan Data

Pada tahap ini, data training dilabeli dengan setiap teks dokumen diberi label positif atau negatif. Proses pelabelan ini dilakukan secara manual.

2.3. Pre-Processing

Tahapan berikutnya adalah pre-processing data. Proses ini bertujuan untuk mengurangi atribut yang tidak diperlukan dalam proses klasifikasi nantinya. Proses *pre-processing* terdiri dari pembersihan (*cleaning*), penyeragaman huruf besar-kecil (*case folding*), penghapusan kata-kata umum (*stopword*), penghapusan kata-kata slang (*slangword*), pemisahan kata (*tokenizing*), dan stemming (mencari kata dasar) terhadap data mentah yang masih kotor.

a. *Cleaning*

Pada prose pertama dari pre-processing yaitu *cleaning*, yang mana dilakukan untuk memeriksa dataset apakah terdapat data yang duplikat, data yang hilang (*missing values*), data yang tidak valid, atau adanya noise pada data. Atribut yang tidak mempengaruhi hasil akan dihapus seperti sentimen, seperti sebutan nama, tanda pagar (*hashtag*), dan tautan (*link*).

Tabel 1. *Cleaning*

Sebelum	Sesudah
Hormati Hasil Pemilu Dengan Damai Dan Aman https://t.co/EvcJfLSw8Q	Hormati Hasil Pemilu Dengan Damai Dan Aman

b. *Case Folding*

Pada tahapan case folding dilakukan untuk mengubah dari huruf kapital pada teks komentar menjadi huruf kecil.

Tabel 2. *Case Folding*

Sebelum	Sesudah
Hormati Hasil Pemilu Dengan Damai Dan Amanh	hormati hasil pemilu dengan damai dan aman

c. *Tokenizing*

Untuk tahapan tokenizing yaitu memisahkan teks menjadi beberapa bagian tersendiri, pada bagian ini digunakan Library Language Toolkit pada google colab phyton.

Tabel 2. *Tokenizing*

Sebelum	Sesudah
hormati hasil pemilu dengan damai dan aman	"hormati" "hasil" "pemilu" "dengan" "damai" "dan" "aman".

d. *Stopword*

Stopword digunakan untuk menghapus kata-kata yang tidak memiliki makna (*stoplist*) atau kata-kata yang tidak mempengaruhi keakuratan proses klasifikasi, seperti kata penghubung. Kata-kata ini akan digantikan dengan spasi dan kata-kata penting akan disimpan dalam *wordlist*.

Tabel 3. *Stopword*

Sebelum	Sesudah
"hormati" "hasil" "pemilu" "dengan" "damai" "dan" "aman".	"hormati" "hasil" "pemilu" "damai" "aman"

e. *Slangword*

Tahapan *slangword* adalah untuk mengganti kata-kata slang atau informal yang sering digunakan dalam kehidupan sehari-hari kedalam bentuk kata baku.

Tabel 4. *Slangword*

Sebelum	Sesudah
tdk	tidak
hati2	hati-hati
yg	yang
ttg	tentang
ngga	tidak

f. *Stemming*

Proses Stemming adalah untuk menghilangkan dan mengembalikan kalimat yang telah diubah ke bentuk aslinya. Penghapusan sufiks dan prefiks menggunakan *library sastrawi*

Tabel 5. *Stemming*

Sebelum	Sesudah
kecurangan pemilu yang terorganisir sejak tingkat pemerintahan tertinggi presiden sampai tingkat terbawah kepala desa baru ada di 2024 pemilu kali ini adalah pemilu terburuk yg pernah terjadi di indonesia	curang pemilu organisir tingkat pemerintahan tinggi presiden tingkat bawah kepala desa baru pemilu buruk indonesia

2.4. Pembobotan Kata

Pada tahapan ini tujuannya untuk meminimalkan data yang tidak relevan sekaligus memaksimalkan data terkait untuk klasifikasi yang akurat. Untuk mencapai hal ini, digunakan teknik *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode ini menghitung kemunculan setiap istilah dalam suatu dokumen dan di seluruh koleksi dokumen.

TF-IDF Merupakan proses untuk menilai seberapa penting sebuah kata (*term*) dalam suatu dokumen dan korpus. Langkah ini dilakukan untuk mengevaluasi bobot atau relevansi kata dalam sebuah dokumen dibandingkan dengan keseluruhan dokumen yang ada dalam korpus (Febriyani & Februariyanti, 2023).

$$IDF(t) = \log \frac{N}{df(t)} \quad (1)$$

Dimana :

TF = Teks frekuensi

IDF = Teks frekuensi dalam dokumen

n = Jumlah dokumen

2.5. Pemodelan Algoritma

Menurut (Murni et al., 2023) Metode Naïve Bayes Classifier untuk analisis sentimen didasarkan pada Teorema Bayes, seperti yang disajikan dalam persamaan berikut :

$$P(C|X) = \frac{P(X|C) * P(C)}{P(X)} \quad (2)$$

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

2.6. Evaluasi

a. Confusion Matrix

Confusion Matrix adalah tabel yang menampilkan empat kombinasi hasil prediksi dan nilai aktual dalam klasifikasi (Ipmawati et al., 2024). Kinerja model dievaluasi menggunakan dua metrik utama *F1-score*, yang menyeimbangkan precision dan recall, serta *accuracy*, yang mengukur ketepatan klasifikasi secara keseluruhan. *F1-score* memberikan gambaran menyeluruh tentang efektivitas model, sedangkan *accuracy* menunjukkan proporsi prediksi yang benar dari total kasus (Murni et al., 2023).

Tabel 6. *Confusion Matrix*

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

Sumber (Murni et al., 2023)

b. K-Fold Cross Validation

K-Fold Cross Validation adalah teknik evaluasi untuk menilai efektivitas model machine learning. Metode ini membagi data menjadi k subset berukuran sama secara acak. Proses validasi dilakukan sebanyak k kali, dengan setiap subset berperan sebagai data uji sekali sementara subset lainnya berfungsi sebagai data latih. Pada setiap iterasi, model dilatih menggunakan data latih dan kinerjanya dievaluasi dengan data uji

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari Twitter (X), pengambilan data dilakukan dengan melakukan crawling (pengambilan data) menggunakan bahasa

pemrograman Python dari Twitter API (Antarmuka Pemrograman Aplikasi) terkait isu kecurangan Pemilu 2024. Pengambilan data dilakukan dari bulan Februari sampai dengan April 2024, dengan jumlah data yang didapatkan adalah sebanyak 4000 data yang dibagi kedalam 5 dataset berdasarkan katakunci pencariannya, dengan masing-masing dataset berjumlah 800 data lebih.

3.2. Pelabelan Data

Setelah data terkumpul, setiap dataset dibagi menjadi dua bagian: 80% digunakan sebagai data latih, sementara 20% sisanya digunakan sebagai data uji. Data latih diberi label secara manual dengan dua kategori: positif dan negatif. Berikut adalah rincian jumlah data berlabel positif dan negatif dalam setiap dataset.

Tabel 7. Jumlah Data Latih

No	Label	Jumlah
1	Pro	258
2	Kontra	542
Jumlah Keseluruhan		800

3.3. Pre-Processing

Proses *Pre-Processing* dimulai dengan tahap *cleaning*, dilanjutkan dengan *case folding*, *tokenizing*, *stopword removal*, penanganan *slangword*, hingga *stemming*. Seluruh proses ini dilakukan menggunakan bahasa pemrograman Python dengan bantuan beberapa library, sebagaimana ditunjukkan pada gambar 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 1. Library dan Source Code Pre-Processing

Hasil dari proses pra-pemrosesan ini akan disimpan dalam file CSV. Berikut adalah data yang telah melalui proses *Pre-Processing* pada salah satu dataset.

Tweet	Label
tidak pernah terjadi sebelumnya kecurangan pemilu yang terorganisir sejak tingkat pemerintahan tertinggi presiden sampai tingkat terbawah kepala desa baru ada di pemilu kali ini adalah pemilu terburuk yg pernah terjadi di Indonesia	1
pemilu paling buruk dalam sejarah kecurangan begitu terstruktur sistematis masif dan brutal	1
mantab prof refly bukti kalau pemilu ini curang makin terlihat mulai dr pencalonan gibran bansos li di tidak mungkin kita permasalahan angka yg didapat dr kecurangan kalau permasalahan berarti mengakui kecurangan itu pemilu ini melanggar konstitusi!	1
kalau mau jujur kecurangan itu sudah terjadi dari putusan mk yang melampaui kewenangan dan terbukti dalam sidang mkmk ada pelanggaran etika berat belum lagi kpu menerima pendaftaran gibran tanpa merubah pkpu mereka pasangan Prabowo gibran saat ini adalah ikon politik	1
kecurangan bukan hanya pada software amp	1
kpu berbuat curang lagi suara diambil dan dimasukan untuk menambah suara dengan cara memalsukan ini petugas kpu memang sudah terangan dan berniat untuk mencurangi kecurangan ini terjadi di tps sumatra utara	1
kecurangan terencana itu dimulai dg penggiringan opini oleh lembaga survei berbayar nanti di hari penghitungan suara lembaga survei berbayar itu akan menyajikan quick count dan hasilnya akan merujuk pd hasil survei sebelumnya masuk akal gak?	1
kecurangan tidak hanya pada salah input data hasil masalah terbesar sebenarnya sebelum pencoblosan keberpihakan kepala negara kepada sang anak mobilisasi kepala daerah aparat desa politik bansos skandal etika di mk penyelenggara pemilu ikut bermain	1
bahkan sesaat sebelum aku menuliskan catatan ini seorang pegawai di tingkat desa menelepon dia dg kesal bercerita bahwa kotak suara hasil pencoblosan yg mestinya wajib disegel saat dibawa ke gudang logistik di batedesa tidak boleh disegel oleh ketuanya alasannya karena mau	1
seperti yang disampaikan ahli tata negara dalam film dokumenter dirty vote desain kecurangan ini tidak muncul semalam tapi sudah lama terstruktur dan sistematis dimulai dari penunjukan pj gubernur dan walikota yang menurut ombudsman ni dinilai maladministrasi	1

Gambar 2. Hasil *Pre-Processing*

3.4. Pembobotan Kata

Penghitungan TF-IDF diimplementasikan dalam kode program Python dengan 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 2. Library *TF-IDF*

3.5. Pemodelan Algoritma

Dataset yang digunakan adalah data training yang sudah melalui tahap pre-processing dan pembobotan TF-IDF. Tahap selanjutnya yaitu proses klasifikasi Menggunakan *naïve bayes multinomial*. Setelah selesai dilatih, model akan disimpan dalam format pickle. Data pickle ini nantinya akan berfungsi sebagai data model untuk melakukan proses klasifikasi pada data testing. Berikut kode program proses klasifikasi Menggunakan *naïve bayes multinomial*

```

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')
    clfm = pickle.load(clf_model)

    y_pred = clfm.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 2. Source Code Naïve Bayes Clasifier

3.6. Evaluasi

a. Confusion Matrix

Hasil dari proses klasifikasi Menggunakan naïve bayes di evaluasi dengan cara menghitung presisi, recall dan F1-score. Proses ini bertujuan untuk mengetahui akurasi performa dari algoritma naïve bayes multinomial pada dataset. Untuk lebih jelasnya lagi, Nilai kinerja untuk setiap label diatas dapat dilihat dalam tabel berikut :

Tabel 8. Hasil *Confusion Matrix*

	Precision	Recall	F1-Score
Pro	0.68	0.99	0.81
Kontra	0.80	0.08	0.14

Tabel diatas menunjukkan bahwa system dapat memperoleh akurasi antara informasi yang tersedia pada system adalah sebesar 68% untuk label pro dan 80% untuk label kontra. System juga memiliki tingkat keberhasilan dalam memulihkan kembali informasi sebesar 8% untuk label kontra dan 99% untuk label pro.

b. K-Fold Cross Validation

Pengujian ulang dilakukan dengan metode K-Fold Cross Validation untuk meningkatkan akurasi dan memastikan evaluasi kinerja algoritma. Pengujian menggunakan *K-fold*

Cross Validation menunjukkan sistem memiliki rata-rata tingkat akurasi pada dataset sebagai berikut :

Tabel 9. Hasil *K-Fold Cross Validation*

	Akurasi	Recall
Pro	78%	100%
Kontra	100%	14%

4. KESIMPULAN

Penerapan metode Naive Bayes dan Cosine Similarity dalam analisis sentiment terhadap Isu Kecurangan Pemilu 2024 dalam platform media sosial Twitter (X) maka dapat disimpulkan sebagai berikut :

1. Tingkat sentimen atau tanggapan pengguna di media sosial X terhadap Isu Kecurangan Pemilu 2024 cenderung negatif. Dapat dilihat dari klasifikasi naive bayes classifier yang menunjukkan presentase sebesar 67,7%, dan presentase akurasi yang ditunjukkan oleh confusion matrik untuk pengujian yaitu sebesar 68% untuk label pro dan 80% untuk label kontra.
2. Selanjutnya, hasil klasifikasi menggunakan Cosine Similarity menunjukkan bahwa kategori dengan bukti kecurangan lebih tinggi dibanding dengan kategori lainnya dengan nilai sebesar 45,5%. Sedangkan untuk kategori lainnya seperti kampanye hitam memiliki persentase sebesar 8,1 %, penggelembungan suara sebesar 12%, manipulasi system sebesar 6,6% dan pelanggaran prosedur sebesar 27,8%.
3. Penerapan K-Fold Cross Validaton menghasilkan nilai akurasi dengan nilai presentase sebesar 78% untuk komentar pro, dan akurasi dengan nilai presentase sebesar 100% untuk komentar kontra.

REFERENSI

- Akbar, Y., & Sugiharto, T. (2023). Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes (Yuma Akbar 1*, Tri Sugiharto 2) Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4.5 dan Naïve Bayes. *Jurnal Sains Dan Teknologi*, 5(1), 115–122. <https://doi.org/10.55338/saintek.v4i3.1368>
- Aryanti, P. G., & Santoso, I. (2023). Analisis Sentimen Pada Twitter Terhadap Mobil Listrik Menggunakan Algoritma Naive Bayes. *IKRA-ITH Informatika: Jurnal Komputer Dan Informatika*, 7(2), 133–137. <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/2821>
- Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1), 34–40. <https://doi.org/10.23960/jitet.v10i1.2262>
- Fauziah, S., Saputra, D. D., Pratiwi, R. L., & Kusumayudha, M. R. (2023). Komparasi Metode Feature Selection Text Mining Pada Permasalahan Klasifikasi Keluhan Pelanggan Industri Telekomunikasi Menggunakan Smote Dan Naïve Bayes. *IJIS - Indonesian Journal On Information System*, 8(2), 174. <https://doi.org/10.36549/ijis.v8i2.289>
- Febriyani, E., & Februariyanti, H. (2023). Analisis sentimen terhadap program kampus merdeka menggunakan algoritma naive bayes classifier di twitter. *Jurnal Tekno Kompak*, 17(1), 25–38. <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/2061>
- Florensius Sianipar, J., Ramadhan, Y. R., & Jaelani, I. (2023). Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes. *Media Online*, 4(1), 360–367. <https://doi.org/10.30865/klik.v4i1.1033>
- Ipmawati, J., Saifulloh, S., & Kusnawi, K. (2024). Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 247–256. <https://doi.org/10.57152/malcom.v4i1.1066>
- Murni, M., Riadi, I., & Fadlil, A. (2023). Analisis Sentimen HateSpeech pada Pengguna Layanan Twitter dengan Metode Naïve Bayes Classifier (NBC). *JURIKOM (Jurnal Riset Komputer)*, 10(2), 566. <https://doi.org/10.30865/jurikom.v10i2.5984>
- Sari, D. N., Sari, D. N., Adelia, F., Rosdiana, F., Butar, B. B., & Hariyanto, M. (2020). Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode

Naive Bayes Classifier. *JIKA (Jurnal Informatika)*, 4(3), 109. <https://doi.org/10.31000/jika.v4i3.3086>

Wulandari, V., Sari, W. J., Alfian, Z., Legito, L., & Arifianto, T. (2024). Implementasi Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor untuk Klasifikasi Penyakit Ginjal Kronik. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 710–718. <https://doi.org/10.57152/malcom.v4i2.1229>