

Naïve Bayes Classifier dan K-Nearest Neighbor pada Analisis Sentimen Perkuliahan Daring di Universitas Muslim Indonesia

Naïve Bayes Classifier and K-Nearest Neighbor on Sentiment Analysis of Online Lectures at Universitas Muslim Indonesia

Silmi Nur Zaskia Wati^{a,1,*}, Herman^{b,2}, Herdianti Darwis^{a,3}

^a Teknik Informatika, Universitas Muslim Indonesia, Jalan Urip Sumoharjo, Makassar, 90231, Indonesia

^b Sistem Informasi, Universitas Muslim Indonesia, Jalan Urip Sumoharjo, Makassar, 90231, Indonesia

¹silminurzaskia@gmail.com; ²herman@umi.ac.id; ³herdianti.darwis@imu.ac.id;

*corresponding author

Informasi Artikel	ABSTRAK
Diserahkan : 10 Maret 2024 Diterima : 29 April 2024 Direvisi : 30 April 2024 Diterbitkan : 30 April 2024 Kata Kunci: Analisis Sentimen Perkuliahan Daring K-Nearest Neighbor Naïve Bayes Classifier Natural Language Toolkit Keywords: Sentiment Analysis Online Learning K-Nearest Neighbor Naïve Bayes Classifier Natural Language Toolkit This is an open access article under the CC-BY-SA license. 	Pelaksanaan perkuliahan daring menggunakan KALAM di UMI banyak menuai kontroversi kalangan mahasiswa. Banyak pendapat mahasiswa yang dikeluarkan terkait metode pembelajaran daring di UMI. Penelitian ini bertujuan menganalisis sentimen mahasiswa terkait perkuliahan daring di UMI dengan menggunakan algoritma <i>Naïve Bayes Classifier</i> dan <i>K-Nearest Neighbor</i>, serta menggabungkan berbagai metode seperti pelabelan dengan NLTK, pengujian dengan 5 <i>cross validation</i>, dan menggunakan <i>unigram tokenizing</i>. Beberapa teknik pelabelan digunakan pada penelitian ini dan menghasilkan tingkat keakuratan paling tinggi adalah pelabelan menggunakan NLTK dengan algoritma KNN dengan menggunakan SMOTE menghasilkan akurasi sebesar 100% dibandingkan dengan algoritma <i>Naïve Bayes Classifier</i> yang memiliki nilai akurasi sebesar 98.33%. Sehingga algoritma KNN dapat digunakan dengan baik pada pengklasifikasian sentimen mahasiswa terhadap perkuliahan daring di UMI. ABSTRACT The implementation of online lectures using KALAM at UMI has generated a lot of controversy among students. Many student opinions were issued regarding online learning methods at UMI. This research aims to analyze student sentiments related to online lectures at UMI by using the <i>Naïve Bayes Classifier</i> and <i>K-Nearest Neighbor</i> algorithms, and combining various methods such as labeling with NLTK, testing with 5 crossvalidation, and using unigram tokenizing. Several labeling techniques are used in this research and the highest accuracy is labeling using NLTK with the KNN algorithm using SMOTE which produces 100% accuracy compared to the <i>Naïve Bayes Classifier</i> algorithm which has an accuracy value of 98.33%. So that the KNN algorithm can be used well in classifying student sentiment towards online lectures at UMI.

I. Pendahuluan

Covid-19 pertama kali ditemukan di Wuhan, Tiongkok pada 31 desember 2019 yang penyebarannya disebabkan oleh Sars-coV2 [1]. Virus Corona telah mewabah di Indonesia sejak bulan Maret 2020. Beberapa kebijakan yang dikeluarkan oleh pemerintah Indonesia untuk mengurangi tingkat penyebaran virus corona dengan memberlakukan *social distancing*, *physical distancing*, hingga pemberlakuan Pembatasan Sosial berskala besar (PSBB) pada beberapa daerah [2]. Di Indonesia, sistem belajar mengajar dilakukan dari rumah, sesuai anjuran pemerintah melalui Menteri Pendidikan dan Kebudayaan (Mendikbud) Indonesia untuk seluruh siswa belajar dari rumah (BDR) [3]. Pandemi Covid-19 memaksa percobaan Pendidikan daring yang hampir belum pernah dilakukan secara serempak sebelumnya pada elemen Pendidikan [4]. Dosen atau pendidik yang merupakan elemen penting dalam pengajaran, diharuskan melakukan alternatif lain dengan beralih dari Pendidikan tradisional (tatap muka) ke pendidikan daring [5].

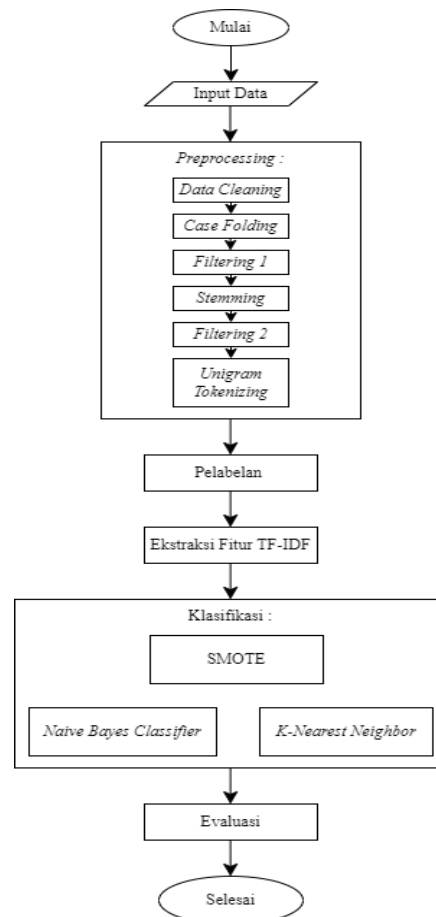
UMI merupakan salah satu universitas yang terdampak wabah Covid-19 yang menyebabkan proses pembelajaran dilaksanakan dalam jaringan atau online. Salah satu media yang digunakan dalam sistem pembelajaran dalam jaringan di UMI yaitu KALAM. Pada 20 Juli 2020, Mendikbud mengeluarkan surat pernyataan tentang keberlanjutan pembelajaran daring meskipun pandemi Covid-19 telah berakhir. Di UMI hingga saat ini masih menerapkan sistem pembelajaran daring menggunakan KALAM yang diterapkan bersamaan dengan pembelajaran tatap muka. Hal tersebut banyak menuai kontroversi kalangan mahasiswa UMI. Banyak pendapat mahasiswa yang dilontarkan terhadap pelaksanaan pembelajaran daring yang bersifat positif, negatif, dan netral.

Analisis sentimen atau biasa di sebut *opinion mining* merupakan proses dalam mengklasifikasikan dokumen teks dengan berlandaskan opini dan pandangan masyarakat yang dilakukan dengan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk menentukan apakah opini tersebut bersifat positif, negatif atau netral. *Naïve Bayes Classifier* adalah salah satu jenis algoritma klasifikasi yang terdapat dalam teknik data mining, algoritma ini digunakan dalam memprediksi sentimen masyarakat dengan cara mengklasifikasi pernyataan yang terkait dengan kasus yang dituju [6]. *Naïve Bayes* menggunakan probabilitas yang sederhana dan berakar pada Teorema Bayes yang mengasumsikan semua atribut independen atau ketidaktergantungan yang tinggi [7]. *Naïve Bayes* menggunakan pendekatan yang sederhana namun kuat dalam melakukan klasifikasi [8]. *K-Nearest Neighbor* merupakan algoritma yang digunakan dalam melakukan proses klasifikasi opini dan pandangan dari hasil analisis sentiment [9], [10].

Penelitian [11] mengklasifikasi tweet mengenai mudik hari raya dengan menggunakan algoritma KNN, SVM, *Naïve Bayes*, dan *Neural Network*. Hasil penelitian ini menunjukkan algoritma KNN memiliki akurasi yang terbaik dibandingkan dengan algoritma lain yaitu sebesar 90,76% dan AUC sebesar 0,939. Penelitian lain yang bertujuan menganalisis opini publik khususnya pada pengguna media sosial mengenai kebijakan pemerintah terkait pembukaan destinasi wisata di masa pandemi Covid-19. Hasil penelitian tersebut menunjukkan nilai tingkat akurasi *Naïve Bayes* tertinggi sebesar 75,53% dibandingkan dengan KNN yang memiliki akurasi sebesar 48,66% [12].

Berdasarkan hasil penelitian tersebut, penelitian ini difokuskan untuk menggunakan algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbor* dalam melakukan analisis sentimen mahasiswa tentang perkuliahan daring di UMI. Keputusan ini didasarkan pada performa algoritma *Naïve Bayes* dan KNN pada penelitian sebelumnya.

II. Metode



Gambar 1. Contoh gambar yang diperbolehkan pada BUSITI

Penelitian ini membandingkan dua algoritma *Naïve Bayes Classifier* dan *K-Nearest Neighbor* dengan menggunakan *library* pelabelan NLTK. Melalui perbandingan algoritma dan pelabelan tersebut diharapkan dapat ditemukan akurasi yang paling optimal dan efektif dalam melakukan klasifikasi.

A. Input Data

Pada penelitian ini dilakukan penginputan data yang telah dikumpulkan melalui kuesioner google form yang disebarakan dilingkungan UML.

B. Preprocessing

Tahap *preprocessing* terdiri dari, pertama *data cleaning* yang bertujuan membersihkan data seperti menghilangkan tanda baca, karakter lain selain alfabet dari A-Z. Kedua, *case folding* yang bertujuan untuk mengubah seluruh huruf dalam teks menjadi huruf kecil. Ketiga, *filtering* menghapus semua kata-kata yang tidak penting dan muncul secara berulang (*stopword*). Keempat, *stemming* bertujuan untuk mencari bentuk kata dasar yang ada dalam teks dengan menggunakan referensi *library* Sastrawi [13]. Kelima, *unigram tokenizing* bertujuan untuk membagi kalimat menjadi token satu kata [14].

C. Pelabelan

Pada penelitian ini dilakukan pelabelan manual oleh responden ketika mengisi kuesioner serta pelabelan menggunakan referensi *library* NLTK untuk menentukan kualitas label yang dihasilkan. Dengan memilih *library* yang memiliki hasil akurasi yang tinggi, diharapkan hasil akhir juga lebih valid. Pelabelan sentimen pada penelitian ini dilabeli dengan sentimen positif, negatif, dan netral.

D. Ekstraksi Fitur

Term Frequency Inverse-Document (TF-IDF) merupakan sebuah metode yang digunakan dalam suatu hubungan antara kata (*term*) dengan sekumpulan dokumen. TF-IDF adalah gabungan dari pendekatan TF dan IDF yang proses perhitungannya mengangkat rasio antara nilai TF dan nilai IDF [13]. Ekstraksi fitur dilakukan dengan menghitung frekuensi kemunculan *term* dalam dokumen dengan menggunakan metode *Term Frequency Inverse Document Frequency* (TF-IDF). Adapun persamaan TF-IDF yaitu [15]:

$$W_{ij} = t_{ij} \cdot idf \quad (1)$$

$$W_{ij} = t_{ij} \cdot \log(idf) \quad (2)$$

Keterangan

W_{ij} : bobot *term* (t_j) terhadap dokumen (d_i)

t_{ij} : banyaknya kemunculan *term* (t_j) dalam (d_i)

n : jumlah semua dokumen dalam *database*

df : jumlah dokumen yang mengandung *term*

E. Klasifikasi

Model klasifikasi pada penelitian ini menggunakan pendekatan *Naïve Bayes Classifier* dan *K-Nearest Neighbor* untuk klasifikasi. *Naïve Bayes Classifier* merupakan salah satu metode *machine learning* yang menerapkan perhitungan probabilitas. *Naïve Bayes* menggunakan probabilitas yang sederhana dan berakar pada Teorema Bayes yang mengasumsikan semua atribut independen atau ketidaktergantungan yang tinggi dari masing-masing kondisi [16]. *K-Nearest Neighbor* (KNN) merupakan proses klasifikasi berdasarkan pada contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, akan tetapi bergantung pada label kategori yang melekat pada dokumen pelatihan yang mirip dengan dokumen teks [17]. Algoritma KNN merupakan metode klasifikasi atas objek yang berdasar pada data training yang jaraknya paling dekat [11].

Pada proses ini, dilakukan pembagian dataset dengan data uji sebesar 80% dan data latih 20% dengan *cross validation* lipatan 5. Selain itu, dilakukan juga teknik *Synthetic Minority Over-sampling* (SMOTE) yang berguna untuk menangani ketidakseimbangan kelas sentimen dalam dataset [18]. Dengan menggunakan teknik ini, kinerja model yang diterapkan dapat meningkatkan akurasi dalam klasifikasi data.

F. Evaluasi

Evaluasi merupakan tahap terakhir dimana performa di ukur berdasarkan akurasi, presisi, dan *recall* dari model yang telah diterapkan. Evaluasi menjadi tahap yang penting untuk mengetahui performa dari algoritma yang diterapkan dalam penelitian ini.

III. Hasil dan Pembahasan

A. Input Data

Dataset pada penelitian ini diambil menggunakan kuesioner yang dibagikan kepada mahasiswa UMI dan disimpan dalam bentuk format *excel*.

B. Preprocessing

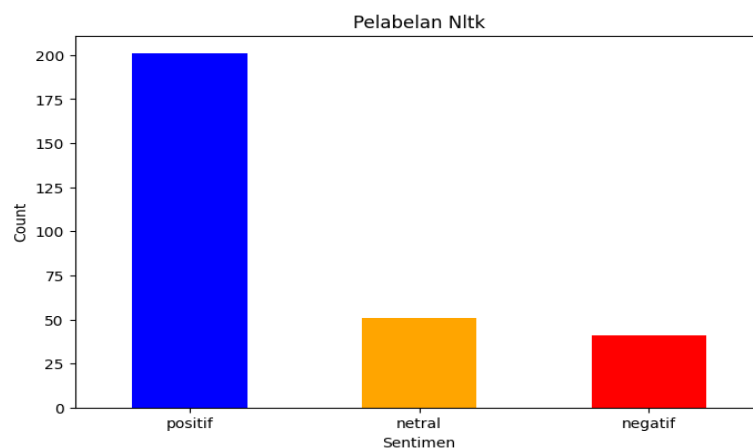
- *Data cleaning* dilakukan untuk menghilangkan tanda baca, karakter lain selain alfabet dari A-Z serta menghapus data kosong dan duplikat.
- *Case folding* dilakukan untuk mengubah seluruh huruf kapitol dalam teks menjadi huruf kecil.
- *Filtering* dilakukan untuk menghapus kata-kata yang tidak penting dan muncul secara rutin (*stopword*) menggunakan referensi dari pujangga serta menghapus kata yang kurang dari 4 huruf.
- *Stemming* dilakukan untuk memperoleh kata dasar dari kata-kata yang ada dalam teks. Proses *stemming* dilakukan dengan menggunakan *library* Sastrawi sebagai referensi.
- *Unigram tokenizing* membagi kalimat menjadi satu kata.

Tabel 1. Hasil Preprocessing

Data Mentah	Data Cleaning	Case Folding	Filtering	Stemming	Tokenizing Unigram
Metode pembelajaran daring menggunakan KALAM UMI memudahkan mahasiswa dalam mengumpulkan tugas secara online. Dosen dapat mengatur dan mengelola tugas dengan mudah, sementara mahasiswa dapat melihat status pengumpulan tugas mereka. Meskipun ada beberapa tantangan, metode ini efektif dalam pengumpulan tugas.	Metode pembelajaran daring menggunakan KALAM UMI memudahkan mahasiswa dalam mengumpulkan tugas secara online. Dosen dapat mengatur dan mengelola tugas dengan mudah sementara mahasiswa dapat melihat status pengumpulan tugas mereka. Meskipun ada beberapa tantangan metode ini efektif dalam pengumpulan tugas	Metode pembelajaran daring menggunakan kalam umi memudahkan mahasiswa dalam mengumpulkan tugas secara online dosen dapat mengatur dan mengelola tugas dengan mudah sementara mahasiswa dapat melihat status pengumpulan tugas mereka meskipun ada beberapa tantangan metode ini efektif dalam pengumpulan tugas	metode pembelajaran daring menggunakan kalam memudahkan mahasiswa dalam mengumpulkan tugas secara online dosen dapat mengatur mengelola tugas dengan mudah sementara mahasiswa dapat melihat status pengumpulan tugas mereka meskipun beberapa tantangan metode efektif dalam pengumpulan tugas	Metode ajar daring guna kalam mudah mahasiswa dalam kumpul tugas cara online dosen dapat atur kelola tugas dengan mudah sementara mahasiswa dapat lihat status kumpul tugas mereka meski beberapa tantang metode efektif dalam kumpul tugas	Metode, ajar, daring, kalam, mudah, mahasiswa, kumpul, tugas, online, dosen, atur, kelola, tugas, mudah, mahasiswa, lihat, status, kumpul, tugas, tantang, metode, efektif, kumpul, tugas

C. Pelabelan

Referensi *library* NLTK menghasilkan akurasi sebesar 98.33% dibandingkan dengan pelabelan manual. Dari 299 dataset, pelabelan menghasilkan 201 sentimen positif, 41 sentimen negatif dan 51 sentimen netral. Hasil pelabelan tersebut nantinya akan diproses untuk klasifikasi *Naïve Bayes* dan KNN.



Gambar 2. Hasil pelabelan NLTK

D. Ekstraksi Fitur TF-IDF

Tahap ini menjelaskan penentuan bobot pada *term*. TF menghitung seberapa sering unigram muncul dalam dokumen, sementara IDF mencerminkan seberapa umum kata tersebut dalam dataset. TF-IDF diperoleh dengan mengalikan nilai TF dan IDF.

Tabel 2. Hasil *Preprocessing*

TERM	TF			DF	n/DF	IDF	W		
	D1	D2	D3				D1	D2	D3
Kalam	1	0	0	1	3	0,47	0,47	0,00	0,00
Fitur	1	0	0	1	3	0,47	0,47	0,00	0,00
Ajar	0	1	0	1	3	0,47	0,00	0,47	0,00
Mahasiswa	1	0	0	1	3	0,47	0,47	0,00	0,00
Mudah	0	1	0	1	3	0,47	0,47	0,47	0,00
Daring	0	1	0	1	3	0,47	0,47	0,47	0,00
Bagus	0	1	0	1	3	0,47	0,47	0,47	0,00
Materi	0	1	0	1	3	0,47	0,47	0,47	0,00
Efektif	0	1	0	1	3	0,47	0,47	0,47	0,47
Kuliah	0	0	1	1	3	0,47	0,47	0,00	0,47
Dosen	0	0	1	1	3	0,47	0,47	0,00	0,00
Metode	0	1	0	1	3	0,47	0,47	0,47	0,00
Paham	1	0	0	1	3	0,47	0,47	0,00	0,00
Sulit	1	0	0	1	3	0,47	0,47	0,00	0,00
Tingkat	1	0	0	1	3	0,47	0,47	0,00	0,00
Susah	0	0	1	1	3	0,47	0,47	0,00	0,47
Cari	0	0	1	1	3	0,47	0,47	0,00	0,47
Kelas	0	0	1	1	3	0,47	0,47	0,00	0,47

E. Evaluasi Hasil Klasifikasi

Performa yang akan diukur yaitu akurasi, presisi dan *recall*. Pada tahap ini, algoritma KNN memiliki tingkat akurasi lebih tinggi dibandingkan dengan algoritma *Naïve Bayes*. Akurasi yang paling tinggi telah melalui penyeimbangan sentimen atau teknik SMOTE.

Tabel 3. Hasil pengujian pada klasifikasi *Naïve Bayes*

	Algoritma	Crossvalidation	Akurasi	Presisi	Recall
Tanpa SMOTE	NBC Pelabelan Manual	5	50.00%	50.00%	100.00%
	NBC Pelabelan NLTK	5	98.33%	100.00%	97.50%
SMOTE	NBC Pelabelan Manual	5	62.50%	76.67%	53.49%
	NBC Pelabelan NLTK	5	98.89%	100.00%	100.00%

Pada Tabel 3, model *Naïve Bayes* menggunakan pelabelan NLTK menghasilkan nilai akurasi terbaik dengan menggunakan teknik SMOTE, dengan 5-*folds cross-validation* menghasilkan nilai akurasi 98.89%.

Tabel 4. Hasil pengujian pada klasifikasi KNN

	Algoritma	Crossvalidation	Akurasi	Presisi	Recall
Tanpa SMOTE	NBC Pelabelan Manual	5	41.67%	50.00%	40.00%
	NBC Pelabelan NLTK	5	96.67%	100.00%	97.50%
SMOTE	NBC Pelabelan Manual	5	62.28%	40.62%	59.09%
	NBC Pelabelan NLTK	5	100.00%	100.00%	100.00%

Pada Tabel 4, model KNN menggunakan pelabelan NLTK menghasilkan nilai akurasi terbaik dengan menggunakan teknik SMOTE, dengan 5-*foldss cross-validation* menghasilkan nilai akurasi 100.00%.

Tabel 5. Tabel *Multiclass Confusion Matrix*

	Predicted Positif	Predicted Netral	Predicted Negatif
Actual Positif	TP	FNeP	FNP
Actual Netral	FNNe	TNe	FPNe
Actual Negatif	FPN	FNeN	TN

Pada Tabel 5, True Positif (TP), True Netral (TNe), True Negatif (TN), False Netral Negatif (FNeN), False Positif Negatif (FPN), False Negatif Netral (FNNe), False Negatif Positif (FNP), False Netral Positif (FNeP).

a) *Naïve Bayes Classifier*

accuracy: 98.33% +/- 3.73% (micro average: 98.28%)

	true Positif	true Netral	true Negatif	class precision
pred. Positif	39	0	0	100.00%
pred. Netral	0	10	0	100.00%
pred. Negatif	1	0	8	88.89%
class recall	97.50%	100.00%	100.00%	

Gambar 3. *Confusion Matrix Naïve Bayes* Pelabelan NLTK

Pada Gambar 3, setelah melalui pengujian terdapat hasil true positif sebanyak 39, yang menunjukkan keberhasilan dalam mengklasifikasikan 39 opini secara benar. Nilai true negatif adalah 8 menandakan bahwa 8 opini negatif berhasil diklasifikasikan dengan tepat. Sementara itu, terdapat 10 opini yang diklasifikasikan sebagai true netral.

$$\text{Akurasi} = \frac{39 + 10 + 8}{39 + 10 + 8 + 1 + 0 + 0 + 0 + 0} \times 100\%$$

$$\text{Akurasi} = \frac{57}{58} \times 100\%$$

$$\text{Akurasi} = 98.33\%$$

Maka didapatkan nilai akurasi sebesar 98.33% dengan perhitungan menggunakan rumus.

$$\text{Presisi} = \frac{39}{39 + 0 + 0} \times 100\%$$

$$\text{Presisi} = \frac{39}{39} \times 100$$

$$\text{Presisi} = 100\%$$

Maka didapatkan nilai presisi sebesar 100% dengan perhitungan menggunakan rumus.

$$\text{Recall} = \frac{39}{39 + 0 + 1} \times 100\%$$

$$\text{Recall} = \frac{39}{40} \times 100\%$$

$$\text{Recall} = 97.50\%$$

Maka didapatkan nilai recall sebesar 97.50% dengan perhitungan menggunakan rumus.

b) *K-Nearest Neighbor*

accuracy: 96.67% +/- 7.45% (micro average: 96.55%)

	true Positif	true Netral	true Negatif	class precision
pred. Positif	39	0	0	100.00%
pred. Netral	1	10	1	83.33%
pred. Negatif	0	0	7	100.00%
class recall	97.50%	100.00%	87.50%	

Gambar 4. *Confusion Matrix KNN* Pelabelan NLTK

Pada Gambar 4, setelah melalui pengujian terdapat hasil true positif sebanyak 39, yang menunjukkan keberhasilannya dalam mengklasifikasikan 39 opini secara benar. Nilai true negatif adalah 7, menandakan bahwa 7 opini negatif berhasil diklasifikasikan dengan tepat. Sementara itu, terdapat 10 opini yang diklasifikasikan sebagai true netral.

- SARS-CoV2 fear on adults' OHRQOL during COVID-19 pandemic," *Saudi journal of Elsevier*, 2021. doi: <http://dx.doi.org/10.1016/j.sjbs.2020.10.033>.
- [5] I. Setyorini, "Pandemi covid-19 dan online learning: apakah berpengaruh terhadap proses pembelajaran pada kurikulum 13?," *J. Ind. Eng. & Management ...*, vol. 1, p. 5, 2020, doi: <https://doi.org/10.30997/karimahtauhid.v1i3.6452>.
- [6] H. Azis, L. Syafie, F. Fattah, and ..., "Unveiling Algorithm Classification Excellence: Exploring Calendula and Coreopsis Flower Datasets with Varied Segmentation Techniques," *2024 18th Int. ...*, 2024, doi: 10.1109/IMCOM60618.2024.10418246.
- [7] I. W. R. Pinastawa and N. A. Arifuddin, "Komparasi Naïve Bayes dan Support Vector Machine dalam Klasifikasi Jenis Citrus.," *Techno. com*, vol. 22, p. 9, 2023, doi: <https://doi.org/10.33633/tc.v22i2.7777>.
- [8] D. Atmajaya, A. Febrianti, and H. Darwis, "Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter," *Indones. J. Comput. ...*, 2023.
- [9] P. R. Prayoga, P. Purnawansyah, T. Hasanuddin, and H. Darwis, "Klasifikasi Daun Herbal Menggunakan K-Nearest Neighbor dan Support Vector Machine dengan Fitur Fourier Descriptor," *Edumatic J. Pendidik. Inform.*, vol. 7, no. 1, pp. 160–168, 2023, doi: 10.29408/edumatic.v7i1.17521.
- [10] I. As'ad, M. A. Asis, H. M. Pakka, R. Mursalin, and Y. binti M. Noor, "K-Nearest Neighbors Analysis for Public Sentiment towards Implementation of Booster Vaccines in Indonesia," *Ilk. J. Ilm.*, vol. 15, no. 2, pp. 365–372, Aug. 2023, doi: 10.33096/ilkom.v15i2.1561.365-372.
- [11] M. Syarifuddin, "ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN," *INTI Nusa Mandiri*, vol. 15, pp. 23–28, Aug. 2020, doi: 10.33480/inti.v15i1.1347.
- [12] D. Era, S. Andryana, and A. Rubhasy, "Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 7, no. 1, pp. 263–272, 2023.
- [13] S. Mutmatimah, Khairunnas, and Khairunnisa, "Metode Deep Learning LSTM dalam Analisis Sentimen Aplikasi PeduliLindungi," *J. Comput. Sci. Informatics*, vol. 1, no. 1, pp. 9–19, 2024, doi: 10.34304/scientific.v1i1.231.
- [14] H. Hayati and M. R. Alifi, "ANALISIS SENTIMEN PADA TWEET TERKAIT VAKSIN COVID-19 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," *JTT (Jurnal Teknol. Ter.)*, vol. 7, p. 110, Oct. 2021, doi: 10.31884/jtt.v7i2.349.
- [15] N. Arifin, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 6, p. 129, Dec. 2021, doi: 10.30998/string.v6i2.10133.
- [16] M. F. Syam, L. N. Hayati, and L. Syafie, "Klasifikasi Pemenuhan Pilar Sanitasi Puskesmas Menggunakan Metode Naive Bayes," *Komputika J. Sist. Komput.*, vol. 12, no. 2, pp. 101–108, 2023, doi: 10.34010/komputika.v12i2.10336.
- [17] D. Zahirah, N. Kurniati, and H. Darwis, "Digital Image Classification of Herbal Leaves Using Knn and Cnn With Glcm Features," *J. Tek. Inform.*, vol. 5, no. 1, pp. 61–67, 2024.
- [18] K. Kim, "Noise Avoidance SMOTE in Ensemble Learning for Imbalanced Data," *IEEE Access*, vol. 9, pp. 143250–143265, 2021, doi: 10.1109/ACCESS.2021.3120738.