

Pemilihan Fitur Chi Square Pada Algoritma Naïve Bayes dan Pengaruhnya Terhadap Analisis Sentimen Masyarakat Indonesia Tentang Pembelajaran Tatap Muka Pada Masa Pandemi Covid-19

A.Habiba^{1*}, R. Rizal Isnanto², J.Endro Suseno³ 

^{1,2,3}Magister Sistem Informasi, Sekolah Pasca Sarjana, Universitas Diponegoro Semarang, Indonesia

ARTICLE INFO

Article history:

Received September 01, 2022

Revised September 03, 2022

Accepted March 21, 2023

Available online April 25, 2023

Kata Kunci:

Analisis Sentimen, Chi Square, Covid-19, Naïve Bayes, Seleksi Fitur.

Keywords:

Sentiment Analysis, Chi Square, Covid-19, Naïve Bayes, Feature Selection.



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2023 by Author. Published by Universitas Pendidikan Ganesha.

ABSTRAK

Kritik dan komentar yang disampaikan masyarakat Indonesia terkait kebijakan pemerintah mengenai pembelajaran tatap muka di masa pandemi Covid-19 menuai pro dan kontra. Tidak sedikit orang tua yang khawatir dengan kebijakan ini dikarenakan para orang tua masih takut akan penyebaran kluster baru Covid-19 di Indonesia yang semakin berkembang, di sisi lain banyak juga orang yang beropini hal ini baik diterapkan mengingat pembelajaran secara daring dinilai kurang efektif karena banyak siswa yang sulit menerima materi yang disampaikan guru secara daring serta banyaknya siswa yang belum memiliki perangkat yang memadai. Banyaknya opini yang dituliskan di Twitter membutuhkan pengklasifikasian sesuai sentimen yang dimiliki agar mudah untuk mendapatkan kecenderungan opini tersebut apakah cenderung beropini netral, positif maupun negatif. Analisis sentimen dalam penelitian ini dilakukan dengan menggunakan metode Naïve Bayes dan seleksi ciri Chi Square dalam melakukan klasifikasi. Hasil analisis yang dari metode Naïve Bayes dengan seleksi ciri Chi Square memiliki akurasi 93% dan tanpa seleksi ciri Chi Square memiliki akurasi 92%, sehingga dapat disimpulkan Metode Naïve Bayes dengan seleksi ciri Chi Square memiliki tingkat akurasi yang lebih baik dibanding tanpa menggunakan seleksi ciri Chi Square.

ABSTRACT

Criticisms and comments submitted to the Indonesian people regarding government policies regarding face-to-face learning during the Covid-19 pandemic have pros and cons. Not a few parents are worried about this policy because parents are still afraid of the spread of the new cluster of Covid-19 in Indonesia which is growing while on the other hand many people who think this is good to apply considering that online learning is considered less effective because many students are difficult to receive material delivered by teachers online and many students who do not have adequate tools. The number of opinions written on Twitter requires classifying according to the sentiments they must get the tendency of these opinions whether they tend to have neutral opinions, positive or negative. Sentiment analysis in this study was conducted using the Naïve Bayes method and chi square's characteristic selection in conducting classification. The results of the analysis of the Naïve Bayes method with chi square characteristic selection have an accuracy of 93% and without chi square characteristic selection has an accuracy of 92%, so it can be concluded the Naïve Bayes Method with chi square characteristic selection has a better accuracy compared without using chi square characteristic selection.

1. PENDAHULUAN

Wabah virus corona atau yang dikenal dengan COVID-19 dinyatakan sebagai pandemi oleh *World Health Organization* (WHO) pada 11 Maret 2020. Akibatnya, banyak negara di dunia menginstruksikan warganya untuk tinggal di rumah, menghindari kontak fisik, dan melakukan pembatasan social (Nahdi, Ramdhani, Yuliatin, & Hadi, 2020; Rizal, Afrianti, & Abdurahman, 2021). Instruksi serupa juga dikeluarkan oleh pemerintah kepada semua institusi pendidikan agar menyelenggarakan kelas secara daring (Fairuz, Ramadhani, & Tanjung, 2021; Sepita & Suryanti, 2020). Maka dari itu, pembelajaran virtual adalah satu-satunya pilihan bagi siswa dan guru untuk berkomunikasi (Widangsa dan Pratama, 2021). Namun semenjak diberlakukannya *new normal*, pembelajaran tatap muka di institusi pendidikan mulai dilakukan pada zona hijau dengan penentuan prioritas berdasarkan jenjang pendidikan yang lebih tinggi terlebih dahulu dan mempertimbangkan kemampuan peserta didik untuk menerapkan protokol kesehatan dan menjaga jarak (Shaleh & Anhusadar, 2021). Presiden Joko Widodo memastikan kegiatan pembelajaran tatap muka secara serempak akan dimulai pada awal September 2021. Hal tersebut dikemukakan Jokowi usai meninjau kegiatan vaksinasi di SMAN 1 Beber, Kabupaten Cirebon, Provinsi Jawa Barat, dikutip

*Corresponding author.

E-mail addresses: Habiba88@gmail.com (A.Habiba)

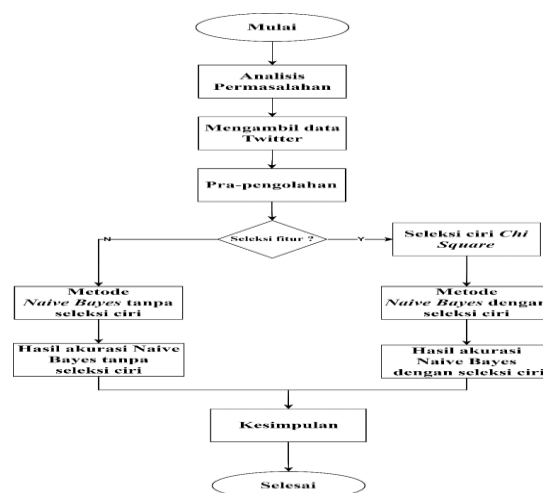
melalui tayangan Youtube Sekretariat Presiden. Saat ini, ada beberapa daerah di wilayah pelaksanaan pemberlakuan pembatasan kegiatan masyarakat (PPKM) level 3 dan 2 yang melakukan uji coba pembelajaran tatap muka. Menurut rencana, jika pembelajaran tatap muka diterapkan secara serempak, maka akan dilakukan maksimal dua kali satu minggu dalam kurun waktu dua jam setiap harinya. Di sela peninjauan vaksinasi, Jokowi berbincang dengan para pelajar yang mengaku sudah tidak sabar untuk melaksanakan kegiatan belajar mengajar tatap muka, sehingga perbincangan ini menjadi *trending* topik di media sosial twitter (Astari, Divayana, & Indrawan, 2020; Fauziyyah, 2020)

Twitter adalah salah satu media pertukaran informasi yang mudah dan populer. Menurut eBizMBA, twitter berada di urutan kedua sebagai media sosial terpopuler setelah Facebook di dalam Top 15 *Most Popular Social Networking Sites* dengan perkiraan pengunjung bulanan yang unik berjumlah 310.000.000 (Berliana, Shaufiah, & Sa'adah, 2018; Rodiyansyah & Winarko, 2012). Twitter merupakan media sosial yang banyak digunakan di Indonesia dan memiliki persebaran serta distribusi informasi yang sangat cepat. Indonesia menempati peringkat 5 pengguna Twitter terbesar di dunia. Oleh karena itu twitter adalah media social yang tepat untuk melakukan penelitian tentang pengklasifikasian opini public (Sari & Yova, 2020; Widya, Prasetya, & Anggrainingsih, 2018). Berdasarkan permasalahan yang terjadi, diperlukan sebuah sistem yang dapat melakukan analisis sentimen terkait pembelajaran tatap muka di masa pandemi. Analisis sentimen dilakukan dengan menggunakan metode *Naïve Bayes* dan seleksi ciri *Chi Square* dalam melakukan klasifikasi (Haryanto, Mawardi, & Muljono, 2018; Kurniawan, 2018). Kelebihan dari *Naïve Bayes* adalah sederhana tetapi memiliki akurasi yang tinggi (Buslim, Putra, & Luh Kesuma Wardhani, 2019; Muhammad, Bukhori, & Pandunata, 2019). Berdasarkan hasil eksperimen, *Naïve Bayes* terbukti dapat digunakan secara efektif untuk mengklasifikasikan berita secara otomatis dengan akurasi mencapai 90.23%. Algoritma *Naïve Bayes* yang sederhana dan kecepatannya yang tinggi dalam proses pelatihan dan klasifikasi membuat algoritma ini menarik untuk digunakan sebagai salah satu metode klasifikasi (Amrullah, Anas, & Hidayat, 2020; Ernawati dkk, 2018).

Pada penelitian ini akan dilakukan penggabungan *Naïve Bayes* dengan seleksi ciri pada Bahasa pemrograman Python. Penyeleksian ciri diperlukan dalam proses memilih subset dari ciri-ciri yang relevan untuk digunakan dalam konstruksi model probabilistik *Naïve Bayes*. Penyeleksian ciri yang digunakan adalah seleksi ciri *chi square* yang bertujuan untuk mengetahui pengaruh terhadap kinerja algoritma *Naïve Bayes* jika menggunakan seleksi ciri *chi square* (Wongkar & Angdresey, 2019). Dari data yang tersedia, sejumlah data akan digunakan untuk menguji hasil klasifikasi sistem *Naïve Bayes* dengan penyeleksian ciri *chi square*. Analisis sentimen yang digunakan bertujuan untuk mengetahui opini publik apakah masyarakat pro dan kontra dengan kebijakan sekolah tatap muka di masa pandemi ini. Hasil analisis ini nantinya dapat digunakan untuk menentukan kebijakan selanjutnya mengenai sekolah tatap muka di masa pandemi Covid-19.

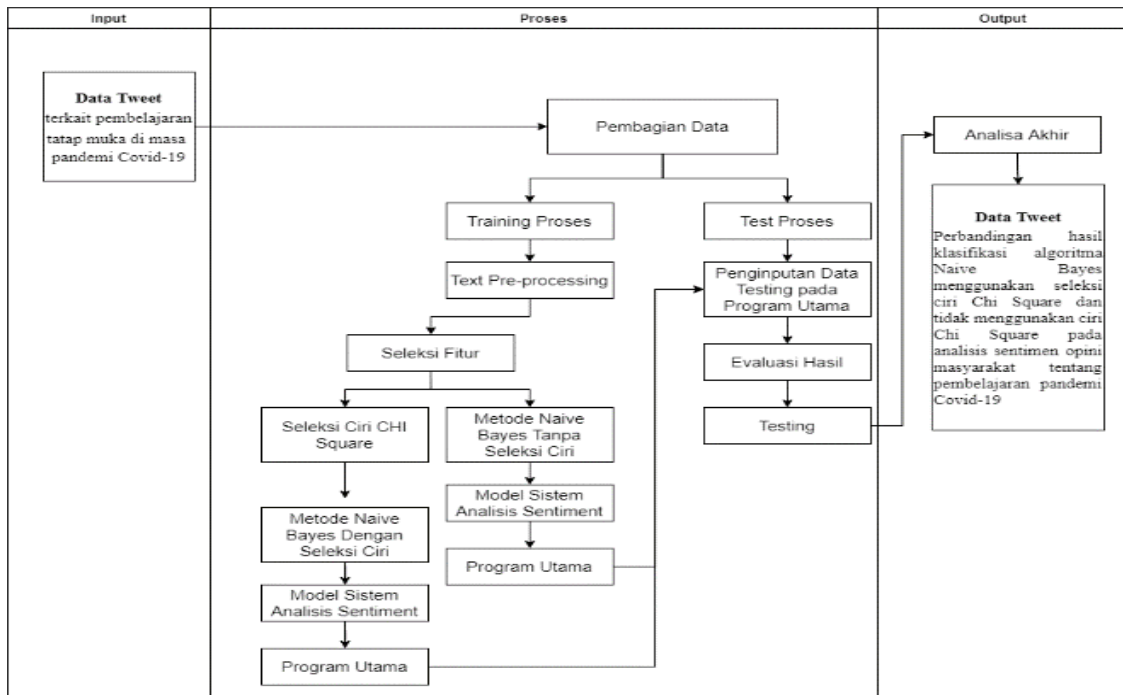
2. METODE

Prosedur penelitian yaitu langkah-langkah yang dipakai untuk mengumpulkan data guna menjawab pertanyaan penelitian yang diajukan di dalam penelitian. Prosedur yang dilakukan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Prosedur Penelitian

Kegiatan pengolahan data terdiri atas 4 tahapan dasar yaitu pengambilan data Twitter (*Web Scraping*), *input*, *proses* dan *output*, seperti pada Gambar 2.



Gambar 2. Kerangka Sistem Informasi

3. HASIL DAN BAHASAN

Hasil

Hasil dari penelitian dibagi menjadi beberapa tahapan. Dalam setiap tahapan akan diberikan penjelasan akan *source code* yang digunakan dan hasil dari tahapan tersebut.

Pengambilan Data Twitter

Pada penelitian ini pengambilan data twitter menggunakan teknik *web scrapping*. *Scraping web* adalah proses menggunakan bot otomatis untuk merayapi internet dan mengekstrak data.

Simpan Data dengan Format CSV

Pada tahap ini untuk menyimpan data dalam *fortmat* dokumen CSV. Peneliti harus mengeskport menggunakan pustaka *pandas* dari *format.txt* kedalam *format csv*. *Source code* dari ekspor ke dalam *format .csv* dapat dilihat pada Gambar 3.

```
import pandas as pd
import sklearn
import numpy as np
import re
import matplotlib.pyplot as plt
import nltk

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import plot_confusion_matrix
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
from sklearn.metrics import f1_score

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
# from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
```

Gambar 3. Source Code Export

Pra Pengolahan

Pada tahap prapengolahan data, data mentah terlebih dahulu dilakukan proses *case folding*, *tokenizing*, *stemming*, serta *filtering*. Hasil dari tahapan ini menghasilkan ciri yang digunakan sebagai data pembelajaran mesin. *Source Code* yang berfungsi untuk *prapengolahan* data seperti *Case Folding*, *tokenizing*, *filtering* dan *stemming* dapat dilihat pada [Gambar 4](#).

	id	date	user	tweet	stem	label
0	1443158819756859904.0	2021-09-29 10:20:49	PiyunganPolsek	Kanit Binmas Iptu Moh. Widayadi Sanani dan Pani...	kanit binmas iptu moh widayadi sanani panit bi...	1
1	1443158763171500032.0	2021-09-29 10:20:36	semarangpedia	Unicef mengapresiasi langkah dan upaya yang di...	unicef apresiasi langkah upaya ambil gubernur ...	1
2	1443158290192430080.0	2021-09-29 10:18:43	PiyunganPolsek	Kanit Binmas Iptu Moh. Widayadi Sanani dan Pani...	kanit binmas iptu moh widayadi sanani panit bi...	1
3	1443157905306360064.0	2021-09-29 10:17:11	JuliyantoNavval	@AndamariUsmani Kepada para kepala sekolah say...	kepala sekolah sila giat ajar tatap muka batas...	1
4	1443157590955889920.0	2021-09-29 10:15:56	hartanto_jhon	@andriawanisa Kepada para kepala sekolah saya ...	kepala sekolah sila giat ajar tatap muka batas...	1
5	1443157566884770048.0	2021-09-29 10:15:51	Spy_Zone85	@ABSetyono @PlateJohnny Setelah Pembelajaran T...	ajar tatap muka ptm sukses jalan prokes sektor...	1
6	1443157253129860096.0	2021-09-29 10:14:36	roysahadoen	RT @Dhinijogja: Presiden @jokowi menginstruksi...	presiden instruksi menkes budi gunadi amp	1
7	1443156016175010048.0	2021-09-29 10:09:41	koranjakarta_id	Menteri Pendidikan, Kebudayaan, Riset, dan Tek...	menteri didik budaya riset teknologi mendikbud...	1
8	1443155396651219968.0	2021-09-29 10:07:13	humasdisdikkab1	Dalam rangka Operasi Patuh Jaya 2021 dan pelak...	rangka operasi patuh jaya laksana ajar tatap m...	-1
9	1443153335951899904.0	2021-09-29 09:59:02	lifeofta	@AREAJULID PTM: Pembelajaran Tatap Muka/njnji...	ptm ajar tatap muka njr gv googling gv kirain...	-1

Gambar 4. Prapengolahan

Seleksi Chi-Square

Proses perhitungan nilai *Chi-Square* dari tiap ciri yang ada pada *Bag of words*. Dengan didapatkannya nilai *Chi-Square* dari tiap kata pada tiap kategori (positif dan negatif), dapat dilakukan eliminasi ciri-ciri berdasarkan rasio jumlah yang diinginkan. Perhitungan ini dapat dilihat pada [Gambar 5](#).

	negative	positive	X(t,c-1) 1	X(t,c-1) 2	X(t,c-1) 3
0	106	456	2566034681515200	705434740990596	3.638
1	4	37	36279853354800	53001707631129	0.685
2	1	31	36291121459200	41387917257216	0.877
3	1	47	87386712883200	62026593716736	1.409
4	267	171	5699291412715200	553696703962596	10.293
...
1707	0	1	43156810800	1295603856009	0.033
1708	0	2	172627243200	2591063748036	0.067
1709	1	0	172886410800	1295603856009	0.133
1710	2	0	691545643200	2591063748036	0.267
1711	1	0	172886410800	1295603856009	0.133

1712 rows x 5 columns

Gambar 5. Hasil Chi Square Negatif dan Positif

Algoritma Naïve Bayes

Pada implementasi algoritma *Naïve Bayes* terdapat dua hasil yang dapat dibandingkan yaitu hasil implementasi metode *Bayes* dengan seleksi ciri *chi Square* dan tanpa seleksi ciri *chi square*.

Confusion Matrix

Pada tahap *Confusion Matrix* dapat ditentukan tingkat akurasi menggunakan tabel hasil pembentukan *confusion matrik* seperti yang disajikan pada [Tabel 1](#) dan [Tabel 2](#).

Tabel 1. *Confusion Matriks Naïve Bayes Tanpa Chi-Square*

		Prediksi	
		Positif	Negatif
Aktual	Positif	196	2
	Negatif	22	81

Tabel 2. *Confusion Matriks Naïve Bayes dengan Chi-Square*

		Prediksi	
		Positif	Negatif
Aktual	Positif	196	2
	Negatif	20	82

Pembahasan

Analisis sentimen dalam penelitian ini diklasifikasikan menjadi sentimen positif dan negatif. Dari 301 data uji yang digunakan didapatkan 198 label sentimen positif atau sebesar 65,78% dan 103 label sentimen negatif atau sebesar 34,22%. Bahwa masyarakat lebih banyak memberikan sentimen positif terkait pembelajaran tatap muka dimasa pandemi, sebagian dari sentiment positif berisi doa dan harapan dalam menyambut pembelajaran tatap muka (Agustin, 2021; Pakpahan & Fitriani, 2020). Namun tidak sedikit juga masyarakat yang memberikan sentimen negatif, beberapa sentimen negatif berisi kekecewaan yang mana salah satu penyebabnya adalah adanya wacana terkait pembelajaran tatap muka di daerah tempat tinggalnya tetapi tidak terjadi (Nissa & Haryanto, 2020; Vhalery, Nur Alfilail, & Robbani, 2021).

4. SIMPULAN

Berdasarkan hasil implementasi dan pengujian pada penelitian ini maka didapatkan kesimpulan sebagai berikut. Dari hasil penelitian yang telah dilakukan didapatkan bahwa sentimen masyarakat terkait pembelajaran tatap muka di tengah pandemi Covid-19 mendapatkan respon yang cenderung positif sebesar 65,78% dan sentimen negatif sebesar 34,22%. Analisis sentimen masyarakat Indonesia tentang pembelajaran tatap muka di masa pandemi Covid-19 dengan metode *Naïve bayes* dan Seleksi ciri *Chi Square* berhasil dijalankan. Hasil analisis yang dari metode *Naïve Bayes* dengan seleksi ciri *Chi Square* memiliki akurasi 93% dan tanpa Seleksi Ciri *Chi Square* memiliki akurasi 92%, Sehingga dapat disimpulkan metode *Naïve Bayes* dengan seleksi ciri *Chi Square* memiliki tingkat akurasi yang lebih baik dibanding tanpa menggunakan seleksi ciri *Chi Square*.

5. DAFTAR PUSTAKA

- Agustin, M. (2021). Tipikal Kendala Guru PAUD dalam Mengajar pada Masa Pandemi Covid 19 dan Implikasinya. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, 5(1), 334–345. <https://doi.org/10.31004/obsesi.v5i1.598>.
- Amrullah, A. Z., Anas, A. S., & Hidayat, M. A. J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal Bumigora Information Technology (BITE)*, 2(1), 40–44. <https://doi.org/10.30812/bite.v2i1.804>.
- Astari, N. M. A. J., Divayana, D. G. H., & Indrawan, G. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1). <https://doi.org/10.30864/jsi.v15i1.332>.
- Berliana, G., Shaufiah, S., & Sa'adah, S. (2018). Klasifikasi Posting Tweet Mengenai Kebijakan Pemerintah Menggunakan Naïve Bayesian Classification. *EProceedings of Engineering*, 5(1).
- Buslim, N., Putra, A. E., & Luh Kesuma Wardhani. (2019). Chi-Square Feature Selection Effect On Naive Bayes Classifier Algorithm Performance For Sentiment Analysis Document. *Conference: 2019 7th International Conference on Cyber and IT Service Management (CITSM)*. <https://doi.org/10.1109/CITSM47753.2019.8965332>.
- Ernawati dkk, S. Y. (2018). Implementation of The Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies. *International Conference on Cyber and IT Service Management (CITSM)*, 1–5. <https://doi.org/10.1109/CITSM.2018.8674286>.
- Fairuz, A. L., Ramadhani, R. D., & Tanjung, N. A. F. (2021). Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter. *Journal of Dinda (Data Science, Information Tehcnology, and Data Analytics)*, 1(1). <https://doi.org/10.20895/dinda.v1i1.180>.

- Fauziyyah, A. K. (2020). Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python. *Jurnal Ilmiah SINUS*, 18(2). <https://doi.org/10.30646/sinus.v18i2.491>.
- Haryanto, A. W., Mawardi, E. K., & Muljono. (2018). Influence of Word Normalization and Chi-Squared Feature Selection on Support Vector Machine (SVM) Text Classification. *International Seminar on Application for Technology of Information and Communication*. <https://doi.org/10.1109/ISEMANTIC.2018.8549748>.
- Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4). <https://doi.org/10.25126/jtiik.201854803>.
- Muhammad, A. N., Bukhori, S., & Pandunata, P. (2019). Sentiment analysis of positive and negative of youtube comments using naïve bayes-support vector machine (nbsvm) classifier. *International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, 199–205. <https://doi.org/10.1109/ICOMITEE.2019.8920923>.
- Nahdi, K., Ramdhani, S., Yuliatin, R. R., & Hadi, Y. A. (2020). Implementasi Pembelajaran pada Masa Lockdown bagi Lembaga PAUD di Kabupaten Lombok Timur. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, 5(1), 177–186. <https://doi.org/10.31004/obsesi.v5i1.529>.
- Nissa, S. F., & Haryanto, A. (2020). Implementasi Pembelajaran Tatap Muka Di Masa Pandemi Covid-19. *Jurnal IKA PGSD (Ikatan Alumni PGSD) UNARS*, 8(2), 402. <https://doi.org/10.36841/pgsdunars.v8i2.840>.
- Pakpahan, R., & Fitriani, Y. (2020). Analisa Pemanfaatan Teknologi Informasi Dalam Pembelajaran Jarak Jauh Di Tengah Pandemi Virus Corona Covid-19. *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, 4(2), 30–36.
- Rizal, M., Afrianti, R., & Abdurahman, I. (2021). Dampak Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) bagi Pelaku Bisnis Coffe shop pada Masa Pandemi Terdampak COVID-19 di Kabupaten Purwakarta The Impact of the Policy for Implementing Community Activity Restrictions for Coffee Shop Busi. *Jurnal Inspirasi*, 12(1), 97–105.
- Rodiyansyah, S. F., & Winarko, E. (2012). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *IJCCS - Indonesian Journal of Computing and Cybernetics Systems*, 6(1). <https://doi.org/10.22146/ijccs.2144>.
- Sari, I. C., & Yova, R. (2020). Sentiment Analysis of the Covid-19 Virus Infection in Indonesian Public Transportation on Twitter Data: A Case Study of Commuter Line Passengers. *International Workshop on Big Data and Information Security (IWBIS)*. <https://doi.org/10.1109/IWBIS50925.2020.9255531>.
- Sepita, S. F., & Suryanti, S. (2020). Pengaruh Pembelajaran Daring Terhadap Hasil Belajar Kognitif Mahasiswa Pada Mata Kuliah Limnologi. *Journal of Research and Education Chemistry*, 2(2), 102. [https://doi.org/10.25299/jrec.2020.vol2\(2\).5826](https://doi.org/10.25299/jrec.2020.vol2(2).5826).
- Vhalery, R., Nur Alfilail, S., & Robbani, H. (2021). Persepsi Mahasiswa Tentang Pembelajaran Online “Google Classroom” Pada Minat Dan Motivasi Belajar. *Intelektium*, 2(1). <https://doi.org/10.37010/int.v2i1.271>.
- Widya, S. S., Prasetya, J. I., & Anggrainingsih, R. (2018). Twitter Sentiment Analysis of Movie Reviews Using Information Gain and Naïve Bayes Classifier. *International Seminar on Application for Technology of Information and Communication*, 190–195. <https://doi.org/10.1109/ISEMANTIC.2018.8549757>.
- Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. *Fourth International Conference on Informatics and Computing (ICIC)*, 1–5. <https://doi.org/10.1109/ICIC47613.2019.8985884>.