

Penelitian

A

Jurnal KRESNA Vol.2 No.2 November 2022

Surat Keputusan Dekan

 UNIVERSITAS BUDI LUHUR FAKULTAS TEKNOLOGI INFORMASI KAMPUS PUSAT : Jl. Raya Gading, Pelukangan Utara, Jakarta Selatan 12280 Telp : (021) 5853753 (Pusat), (021) 7371964, Fax : (021) 5853753 Website : http://www.budiluhur.ac.id		 UNIVERSITAS BUDI LUHUR FAKULTAS TEKNOLOGI INFORMASI KAMPUS PUSAT : Jl. Raya Gading, Pelukangan Utara, Jakarta Selatan 12280 Telp : (021) 5853753 (Pusat), (021) 7371964, Fax : (021) 5853753 Website : http://www.budiluhur.ac.id	
KEPUTUSAN DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR			
NOMOR : K/UBL/FTI/000/004/06/22			
TENTANG: PENUGASAN KEGIATAN TRI DHARMA & PENUNJANG BAGI DOSEN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR SEMESTER GASAL TAHUN AKADEMIK 2022/2023			
DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR			
Menimbang	1) Bahwa Dosen adalah pendidik profesional dan ilmu dengan tugas utama mentransformasikan, mengembangkan, dan menyebarkan ilmu pengetahuan, teknologi, dan seni melalui pendidikan/pengajaran penelitian & karya ilmiah, dan Pengabdian pada masyarakat yang dikenal dengan istilah Tri Dharma Perguruan Tinggi; 2) Bahwa untuk meningkatkan profesionalitas dan kompetensi sebagai pendidik profesional maka dipandang perlu untuk memberikan tugas-tugas tambahan/penunjang dalam lingkup kegiatan penunjang Tri Dharma;	KEDUA	d. Pengabdian Kepada Masyarakat (PPM), dalam kegiatan terprogram, terjadwal atau insidental;
Mengingat	1) Undang – undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi; 2) Peraturan Pemerintah Republik Indonesia Nomor 17 Tahun 2010 tentang Pengelolaan dan Penyelenggaraan Pendidikan; 3) Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia Nomor 44 Tahun 2015 tentang Standar Nasional Pendidikan Tinggi; 4) Peraturan Menteri Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia Nomor 15 Tahun 2017 tentang Penamaan Program Studi Pada Perguruan Tinggi; 5) Akta Yayasan Pendidikan Budi Luhur Tanggal 23 Desember 1991; 6) Peraturan Pengurus Yayasan Pendidikan Budi Luhur Cakdi Nomor: K/YBL/KEP/000/389/06/17 tanggal 24 Agustus 2017 tentang Statuta Universitas Budi Luhur;	KETIGA	d. Dosen-dosen yang melaksanakan penugasan wajib membuat Laporan Kegiatan, dengan mengikuti pedoman dari Fakultas/Program Studi, sebagai pertanggungjawaban atas kegiatan yang diikuti;
Menetapkan	MEMUTUSKAN 1) Menugaskan dosen-dosen Fakultas Teknologi Informasi Universitas Budi Luhur untuk melaksanakan kegiatan Tri Dharma Perguruan Tinggi dan penunjangnya pada Semester Gasal Tahun Akademik 2022/2023 yang meliputi: a. Kegiatan partisipasi aktif dalam Pertemuan Ilmiah sebagai Ketua/Anggota/Peserta/Pembicara/Penulis/Narasumber pada kegiatan Seminar, Workshop, Konferensi, Pelatihan, Simposium, Lokakarya, Forum Diskusi, Sosialisasi dan sejenisnya; b. Publikasi Ilmiah pada Prosiding, Jurnal/majalah/jurnal, buku dan sejenisnya; c. Partisipasi dalam organisasi profesi, organisasi keilmuan dan/atau organisasi lain yang menunjang kegiatan Tri Dharma Pendidikan Tinggi;	KEEMPAT	d. Kegiatan Tri Dharma yang tidak termasuk dalam surat keputusan ini akan memiliki penugasan tersendiri; d. Keputusan ini berlaku sejak tanggal ditetapkan dan akan diubah sebagaimana mestinya apabila di kemudian hari terdapat kekeliruan.
		Ditetapkan di : Jakarta Pada Tanggal : 1 September 2022  Dr. Ir. Deni Mahdiana, S.Kom, M.M., N.Kom	



Samsinar Samsinar

kepada Andreansyah, saya ▾

11.50 (14 menit yang lalu)



Andreansyah Andreansyah, Grace Gata:

Naskah anda berjudul, "Analisis Sentimen Untuk Program Vaksin Booster Sebagai Syarat Mudik 2022 Menggunakan Algoritma KNN," terpilih untuk diterbitkan pada Jurnal Kresna Vol.2 No.2 November 2022.

...

↩ Balas

↩ Balas ke semua

➡ Teruskan

<https://jurnaldrpm.budiluhur.ac.id/index.php/Kresna/issue/current>

jurnaldrpm.budiluhur.ac.id/index.php/Kresna

KRESNA: Jurnal Riset dan Pengabdian Masyarakat

[Home](#) [Current](#) [Archives](#) [Announcements](#) [About ▾](#)

About the Journal

Jurnal KRESNA adalah media publikasi ilmiah online yang dikelola oleh Direktorat Riset dan Pengabdian Kepada Masyarakat Universitas Budi Luhur. Jurnal KRESNA diterbitkan secara berkala oleh DRPM Universitas Budi Luhur. Jurnal KRESNA mulai terbit pada bulan November tahun 2021, mulai tahun 2022 frekuensi terbitan menjadi 2 (dua) kali dalam setahun, yaitu bulan Mei dan November. Jurnal ini mendapatkan E-ISSN 2809-6509 Mulai Edisi Vol. 1, No. 1, November 2021.

Karya ilmiah yang masuk akan melalui proses review oleh tim reviewer.

Current Issue

Vol. 2 No. 2 (2022): November



Title : KRESNA: Jurnal Riset dan Pengabdian Kepada Masyarakat

Publication Schedule : Mei, November

ISSN : 2809-6509 (Online)

Accreditation : -

Subject : Multidisiplin Ilmu (Hasil Riset dan Pengabdian Kepada Masyarakat)

Indexed at : [Google Scholar](#), [Portal Garuda](#)

DOI : 10.36080

Publisher : DRPM Universitas Budi Luhur

Pada issue ini topik yang diterbitkan sebanyak 10 artikel dengan berbagai topik Multidisiplin Ilmu dari Hasil Riset dan Pengabdian Kepada Masyarakat yang relevan dalam bidang Computer Science and Electrical Engineering, Business and Economics, Communication and Creative Design, Social and Political Science dan Architecture.

Cover, Kata Pengantar, Daftar Isi [Download](#)

Faktor yang Mempengaruhi Audit Delay Perusahaan Sektor Property dan Real Estate di BEI Periode 2015-2020

Factors Influencing Audit Delay for Companies in the Property and Real Estate Sector on the IDX for the 2015-2020

Resa Sekar Putri, Martini Martini

175-184

**Bank Indonesia bersama Perbankan Syariah Membantu Digitalisasi UMKM Dengan Manfaatkan Teknologi Quick Response Code Indonesia Standard QRIS**

Bank Indonesia and Islamic Banking Assist in the Digitization of UMKM by Utilizing Quick Response Code Indonesia Standard (QRIS) Technology

Didik Hariyadi Raharjo, Selamat Riyadi

185-192

**Analisis Sentimen untuk Program Vaksin Booster Sebagai Syarat Mudik 2022 Menggunakan Algoritma KNN**

Sentiment Analysis for The Booster Vaccine Program as A Condition for Homecoming 2022 Using The KNN Algorithm

Andreansyah, Grace Gata

193-201

**Implementasi Web Service Untuk Manajemen Akun Linux Pada Server Badan Litbang Perhubungan Menggunakan Php**

Implementation of Web Service for Linux Account Management On Badan Litbang Perhubungan Using Php

Dewi Wachyuni, Joko Christian Candra

202-211

**Implementasi Makopala Network Server Pada Router Mikrotik Sebagai Aplikasi Usermanager Untuk Kampung Wifi Berbasis Web**

Implementasi Makopala Network Server Pada Router Mikrotik Sebagai Aplikasi Usermanager Untuk Kampung Wifi Berbasis Web

Yuga Prasetyo, Hari Soetanto

212-220



1



2



KRESNA: Jurnal Riset dan Pengabdian Masyarakat
Volume 2, Nomor 2, November, 2022, Hal: 193-201

E-ISSN: 2809-6509

Analisis Sentimen untuk Program Vaksin Booster Sebagai Syarat Mudik 2022 Menggunakan Algoritma KNN

Sentiment Analysis for The Booster Vaccine Program as A Condition for Homecoming 2022 Using The KNN Algorithm

Andrianyah^{1*}, Grace Gata²

¹Takutan Teknologi Informasi
Universitas Budi Luhur
Email: ¹andrianyah27@gmail.com, ²grace.gata@budiuhur.ac.id
(* - corresponding author)

Abstract
Social media as a place to access and disseminate information has increased, one of which is Twitter. Twitter in this study is used as a source to find public opinion on government policies regarding booster vaccines as a condition for going home in Indonesia in 2022. These opinions can be positive or negative, depending on the public's view of the object. So the purpose of this study is to conduct a sentiment analysis using the K-Nearest Neighbors algorithm to determine how much accuracy is generated from the K-Nearest Neighbors algorithm in classifying the sentiment of a tweet related to booster vaccines as a condition for going home. This study uses the Cross-Industry Standard Process For Data Mining (CRISP-DM) methodology. By going through several stages, such as preprocessing and modeling steps. The data obtained is 2,344 with positive sentiment 1,070 tweets and negative views 414 tweets. Based on the results obtained from the modeling stage using the K-Nearest Neighbors algorithm for split data modeling comparison 70:30 using k=9, the accuracy obtained is 85.17%, precision is 85.51%, and recall is 98.82%.

Keywords : sentiment analysis, text mining, k-nearest neighbors, homecoming, vaccine booster.

Abstrak
Media sosial sebagai tempat untuk mengakses dan menyebarkan informasi telah berkembang sangat pesat, salah satunya adalah *twitter*. *Twitter* pada penelitian ini digunakan sebagai sumber untuk mengetahui pendapat atau opini masyarakat terhadap kebijakan pemerintah tentang vaksin *booster* sebagai syarat mudik di Indonesia tahun 2022. Opini-opini tersebut bisa berupa opini positif atau negatif, tergantung dari pandangan publik terhadap objek tersebut. Maka tujuan dari penelitian ini adalah untuk melakukan analisis sentimen dengan menggunakan algoritma *K-Nearest-Neighbors*, untuk mengetahui seberapa besar akurasi yang dihasilkan dari algoritma *K-Nearest Neighbors* dalam mengklasifikasi sentimen dari sebuah *tweet* yang berkaitan dengan vaksin *booster* sebagai syarat mudik. Pada penelitian ini menggunakan metodologi *Cross-Industry Standard Process For Data Mining (CRISP-DM)*. Dengan melewati beberapa tahapan seperti tahapan *preprocessing* dan pemodelan. Jumlah data yang didapat 2344 dengan sentimen positif 1070 *tweet* dan sentimen negatif 414 *tweet*. Berdasarkan hasil yang didapat dari tahapan pemodelan dengan menggunakan algoritma *K-Nearest Neighbors* pemodelan *split data* perbandingan 70:30 menggunakan k=9, maka *accuracy* yang dihasilkan sebesar 85.17%, *precision* 85.51% dan *recall* 98.82%.

Kata Kunci : analisis sentimen, *text mining*, *k-nearest neighbors*, mudik, vaksin *booster*.

1. PENDAHULUAN
Wabah *Covid-19* ditetapkan secara resmi sebagai pandemi global oleh *World Health Organization (WHO)* pada tanggal 11 Maret 2020. Penyebaran virus tersebut dimulai pada akhir tahun 2019. Masa Pandemi *Covid-19* di tahun ke-3 saat ini masih terus dilakukan walaupun bagi seluruh masyarakat [1].

<https://jurnalrpm.budiuhur.ac.id/index.php/Kresna/>

193

1



2



KRESNA: Jurnal Riset dan Pengabdian Masyarakat
Volume 2, Nomor 2, November, 2022, Hal: 193-201

E-ISSN: 2809-6509

Analisis Sentimen untuk Program Vaksin Booster Sebagai Syarat Mudik 2022 Menggunakan Algoritma KNN

Sentiment Analysis for The Booster Vaccine Program as A Condition for Homecoming 2022 Using The KNN Algorithm

Andrianyah^{1*}, Grace Gata²

¹Takutan Teknologi Informasi
Universitas Budi Luhur
Email: ¹andrianyah27@gmail.com, ²grace.gata@budiuhur.ac.id
(* - corresponding author)

Abstract
Social media as a place to access and disseminate information has increased, one of which is Twitter. Twitter in this study is used as a source to find public opinion on government policies regarding booster vaccines as a condition for going home in Indonesia in 2022. These opinions can be positive or negative, depending on the public's view of the object. So the purpose of this study is to conduct a sentiment analysis using the K-Nearest Neighbors algorithm to determine how much accuracy is generated from the K-Nearest Neighbors algorithm in classifying the sentiment of a tweet related to booster vaccines as a condition for going home. This study uses the Cross-Industry Standard Process For Data Mining (CRISP-DM) methodology. By going through several stages, such as preprocessing and modeling steps. The data obtained is 2,344 with positive sentiment 1,070 tweets and negative views 414 tweets. Based on the results obtained from the modeling stage using the K-Nearest Neighbors algorithm for split data modeling comparison 70:30 using k=9, the accuracy obtained is 85.17%, precision is 85.51%, and recall is 98.82%.

Keywords : sentiment analysis, text mining, k-nearest neighbors, homecoming, vaccine booster.

Abstrak
Media sosial sebagai tempat untuk mengakses dan menyebarkan informasi telah berkembang sangat pesat, salah satunya adalah *twitter*. *Twitter* pada penelitian ini digunakan sebagai sumber untuk mengetahui pendapat atau opini masyarakat terhadap kebijakan pemerintah tentang vaksin *booster* sebagai syarat mudik di Indonesia tahun 2022. Opini-opini tersebut bisa berupa opini positif atau negatif, tergantung dari pandangan publik terhadap objek tersebut. Maka tujuan dari penelitian ini adalah untuk melakukan analisis sentimen dengan menggunakan algoritma *K-Nearest-Neighbors*, untuk mengetahui seberapa besar akurasi yang dihasilkan dari algoritma *K-Nearest Neighbors* dalam mengklasifikasi sentimen dari sebuah *tweet* yang berkaitan dengan vaksin *booster* sebagai syarat mudik. Pada penelitian ini menggunakan metodologi *Cross-Industry Standard Process For Data Mining (CRISP-DM)*. Dengan melewati beberapa tahapan seperti tahapan *preprocessing* dan pemodelan. Jumlah data yang didapat 2344 dengan sentimen positif 1070 *tweet* dan sentimen negatif 414 *tweet*. Berdasarkan hasil yang didapat dari tahapan pemodelan dengan menggunakan algoritma *K-Nearest Neighbors* pemodelan *split data* perbandingan 70:30 menggunakan k=9, maka *accuracy* yang dihasilkan sebesar 85.17%, *precision* 85.51% dan *recall* 98.82%.

Kata Kunci : analisis sentimen, *text mining*, *k-nearest neighbors*, mudik, vaksin *booster*.

1. PENDAHULUAN
Wabah *Covid-19* ditetapkan secara resmi sebagai pandemi global oleh *World Health Organization (WHO)* pada tanggal 11 Maret 2020. Penyebaran virus tersebut dimulai pada akhir tahun 2019. Masa Pandemi *Covid-19* di tahun ke-3 saat ini masih terus dilakukan walaupun bagi seluruh masyarakat [1].

Melihat pesatnya penyebaran *Covid-19* dan dampak yang akan ditimbulkan apabila tidak secepatnya dicegah, salah satu solusi untuk mengurangi tingkat penyebaran virus yaitu dengan Vaksin *Covid-19*. Vaksin *Covid-19* melalui program vaksin dosis 1 dan 2, serta saat ini terus dilakukan pelayanan pemberian Vaksin *Booster Covid-19* ketiga yang bisa didapatkan 6 (enam) bulan setelah dilakukan vaksin kedua. Vaksin akan melindungi masyarakat terhadap penyebaran virus dan juga dapat menurunkan tingkat penyebaran virus tersebut [2].

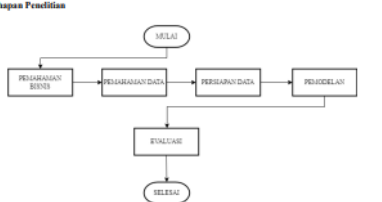
Vaksin *Booster Covid-19* ketiga juga menjadi salah satu syarat mudik di tahun 2022. Mudik adalah kebiasaan masyarakat yang biasa dilakukan pada saat menjelang Idul Fitri. Mudik juga merupakan kegiatan perjalanan dari kota ke desa yang dilakukan secara serentak atau massal, selain dari pertimbangan budaya dan agama [3].

Namun banyak masyarakat yang setuju dan tidak setuju terhadap Vaksin *Booster Covid-19* ketiga sebagai syarat mudik 2022. Sehingga hal tersebut menjadi pembicaraan di berbagai media sosial, salah satunya media sosial *Twitter*. Opini masyarakat di media sosial *Twitter* pun sangat beragam, sehingga terdapat respon yang berbeda dari setiap masyarakat. Kemungkinan ada respon dari masyarakat yaitu respon positif maupun negatif [4]. Respon masyarakat dapat dilihat pada beberapa *tweet* dengan menggunakan keyword "Vaksin *Booster*" dan "Mudik". Sehingga, dibutuhkan penelitian sentimen untuk mengetahui opini-opini tersebut dengan tujuan dijadikan referensi bagi pemerintah atau pihak berkepentingan lainnya dalam membuat peraturan mudik di tahun 2022 atau tahun berikutnya [5].

Penelitian ini menggunakan metode *K-Nearest Neighbors (KNN)*. KNN adalah salah satu algoritma yang mudah diimplementasikan dengan tingkat keefektifan yang tinggi, serta cocok untuk berbagai masalah yang berhubungan dengan klasifikasi [6]. Maka tujuan dari penelitian ini adalah melakukan analisis sentimen dengan menggunakan algoritma KNN dalam mengklasifikasi sentimen dari sebuah *tweet* yang berkaitan dengan vaksin *booster* sebagai syarat mudik 2022.

2. METODE PENELITIAN

2.1 Tahapan Penelitian



Gambar 1: Tahapan Penelitian Analisis Sentimen untuk Program Vaksin Booster Sebagai Syarat Mudik 2022 menggunakan Algoritma KNN

Pada penelitian ini terdiri dari beberapa tahapan yang ditunjukkan pada gambar 1. Penelitian ini menerapkan metodologi *Cross-Industry Standard Process For Data Mining (CRISP-DM)*. Berikut adalah penjelasan mengenai tahapan-tahapan yang dilakukan selama penelitian [7].

a. Pemahaman Bisnis
Pada tahapan ini peneliti memahami permasalahan yang diangkanya yaitu analisis sentimen terhadap vaksin *booster* sebagai syarat mudik 2022, penelitian ini dilakukan dengan mempelajari artikel atau berita di internet tentang vaksin *booster* sebagai syarat mudik 2022, kemudian menggali informasi

<https://jurnalrpm.budiuhur.ac.id/index.php/Kresna/>

194



1



2



3

KRESNA: Jurnal Riset dan Pengabdian Masyarakat
Volume 2, Nomor 2, November, 2022, Hal: 193-201 E-ISSN: 2809-6509

dan mencari data tweet yang bersumber dari media sosial twitter yang berkaitan dengan vaksin booster dan mudik, dari postingan tweet tersebut bisa mengetahui komentar dari masyarakat.

b. Pemahaman Data

Pada tahapan ini peneliti memproses pengambilan data melalui twitter dengan menggunakan keyword seperti vaksin booster dan mudik. Proses pengambilan data pada penelitian ini menggunakan *software rapid minor*. Sebelum melakukan proses pengambilan *tweet*, peneliti diwajibkan untuk membuat koneksi yang menghubungkan antara *rapid minor* dengan *twitter* agar proses pengumpulan data *tweet* dapat dilakukan. Data yang dikumpulkan dari tanggal 27 Mei 2022 hingga 30 Mei 2022. Dengan data *tweet* yang dikumpulkan sebanyak 8341 *tweet*. Data yang diolah adalah data yang sesuai dengan keyword sebelumnya yang telah disesuaikan. Setelah mengumpulkan data yang berkaitan dengan vaksin booster dan mudik, kemudian data digabung menjadi satu, lalu peneliti melakukan *remove duplicate* menggunakan *rapid minor* untuk menghapus data yang muncul secara berulang. Setelah melakukan *remove duplicate* maka sebelum melakukan *preprocessing* data, data perlu diberi label sentimen terlebih dahulu, sebelum data dilakukan dengan menggunakan metode *lexicon*.

c. Persiapan Data

Pada tahapan ini data yang sudah didapatkan sebelumnya, akan ditentukan sentimenya. Setelah pelabelan sentimen dilakukan, maka data *tweet* yang sudah memiliki label sentimen akan dilakukan data *preprocessing*. Tahapan-tahapan saat melakukan *preprocessing* [8]:

- 1) *Remove Duplicate* adalah proses untuk menghapus data yang muncul secara berulang.
- 2) *Replace* adalah proses untuk menghilangkan *stopwords* yang terdapat pada *tweet* seperti RT, https, dan @username.
- 3) *Case Folding* adalah operator yang berfungsi untuk mengubah semua huruf menjadi huruf kecil.
- 4) *Tokenizing* adalah proses untuk memecah data *tweet* yang masih dalam bentuk kalimat menjadi individual.
- 5) *Filter Tokens (By Length)* adalah operator untuk memfilter kata berdasarkan panjang karakter.
- 6) *Stem (Dictionary)* adalah proses mengembalikan kata menjadi bentuk kata dasar dari sebuah kata.
- 7) *Stopwords (Dictionary)* adalah proses yang dilakukan untuk menghilangkan kata sambungan.
- 8) *Term Weighting (TF-IDF)* adalah proses untuk melakukan pembobotan kata, dengan cara mencari nilai *Term Frequency* (TF), lalu mencari nilai *Document Frequency* (DF), kemudian mencari nilai *Inverse Document Frequency* (IDF), dan menghitung bobot [9].

d. Pemodelan

Pada tahapan ini peneliti akan melakukan pemodelan dengan menggunakan data tahapan yaitu split data dan *cross fold validation*. Pada split data membagi dataset tersebut dengan perbandingan data *training* dan data *testing* yaitu 60:40, 70:30, dan 80:20. Lalu pada tahap *Cross Validation* dengan pembagian secara acak kedalam 10 bagian (*number of fold* =10). Pada tahap pemodelan ini menggunakan algoritma KNN namun sebelumnya telah dilakukan perbandingan model klasifikasi dengan menggunakan algoritma *Naive Bayes* dan SVM [10].

e. Evaluasi

Pada tahapan ini peneliti akan melakukan evaluasi metode klasifikasi dengan mengukur performa menggunakan *confusion matrix* terhadap algoritma *K-Nearest Neighbors*.

3. HASIL DAN PEMBAHASAN

3.1 Pemahaman Terhadap Data

Pada tahapan ini peneliti melakukan pengumpulan data dari media sosial *twitter* dengan menggunakan *rapidminer* serta peneliti akan memahami data yang akan digunakan, kemudian peneliti menentukan keyword yang berkaitan dengan vaksin booster sebagai syarat mudik 2022 untuk proses *crawling data tweet*. Peneliti juga melakukan pengambilan data pada tanggal 27 Mei 2022 hingga 30 Mei 2022. Berikut ini adalah data *tweet* yang berhasil peneliti kumpulkan dari *twitter* dengan memakai *web rapidminer* seperti yang ditunjukkan pada Tabel 1.

<https://jurnalidpn.badiluhur.ac.id/index.php/Kresna/> 195



2



3



4

KRESNA: Jurnal Riset dan Pengabdian Masyarakat
Volume 2, Nomor 2, November, 2022, Hal: 193-201 E-ISSN: 2809-6509

Tabel 1. Jumlah Data *Post Hasil Crawling*

Keyword	Tanggal	Jumlah
Vaksin Booster	27 - 30 Mei 2022	4536
Mudik	27 - 30 Mei 2022	3805
Jumlah		8341

3.2 Persiapan Data

Pada tahapan persiapan data ini peneliti akan membagi menjadi 2 tahapan. Pada tahapan *preprocessing* proses yang akan dilakukan oleh peneliti setelah data *tweet* sudah memiliki sentimen positif atau negatif dan tahapan penentuan data *latih* dan data *uji*, tahapan ini merupakan proses menentukan perbandingan data *latih* dan data *uji* terlebih dahulu sebelum masuk ke dalam tahapan pemodelan.

3.2.1 Tahapan Preprocessing

Menyediakan data dilakukan dengan menggunakan tahapan *preprocessing* yang terdiri dari *tokenize*, *transform case*, *filter tokens (by Length)*, *stem (Dictionary)* dan *filter stopwords (Dictionary)*. Dimana dalam penggunaan *filter stopwords* peneliti menggunakan kamus *stopwords* milik *Kaggle.com* dan penggunaan *stem* peneliti membuat manual dengan bantuan kamus besar bahasa Indonesia. Simpan kata-kata yang ada dalam dataset dalam bentuk *file notepad* atau *.txt* yang memproses dataset dengan *rapid minor* yang selanjutnya membuat kalimat menjadi kata dasar. Berikut ini adalah contoh hasil dari tahapan *preprocessing* Tabel 2.

Tabel 2. Perbandingan Sebelum dan Setelah Preprocessing

Preprocessing	Sebelum	Setelah
Transform Case	Presiden Jokowi Minta Masyarakat Tak Pindah Jera Vaksin Booster	presiden jokowi minta masyarakat tak pindah jera vaksin booster
Tokenize	Jakarta Perkaras Suk Vaksin Booster Covid RI Lebih dan Cukup	Jakarta Perkaras Suk Vaksin Booster Covid RI Lebih dan Cukup
Filter Tokens (by Length)	Akhiradilalah di mudik kemari berlangsung aman dan saat tanpa ada peristiwa yang berakibat fatal serta tidak ada penemuan kasus Covid19 yang berarti apa Presiden	Akhiradilalah mudik kemari berlangsung aman dan saat tanpa ada peristiwa yang berakibat fatal serta tidak penemuan kasus Covid19 berarti apa presiden
Filter Stopwords (Dictionary)	Akhiradilalah mudik kemari berlangsung aman dan saat tanpa ada peristiwa yang berakibat fatal serta tidak penemuan kasus Covid19 berarti presiden	Akhiradilalah mudik kemari aman dan saat tanpa ada peristiwa yang berakibat fatal serta penemuan Covid19 presiden
Stem (Dictionary)	Akhiradilalah mudik kemari saat tanpa peristiwa berakibat fatal serta penemuan Covid19 presiden	Akhiradilalah mudik kemari saat tanpa peristiwa akibat fatal serta tambah Covid19 presiden

3.2.2 Pembagian Data Training dan Data Testing

Data *latih* adalah data yang akan digunakan untuk proses *latih* bagi model klasifikasi. Sedangkan data *uji* adalah data yang akan digunakan untuk uji *rule* klasifikasi. Pada penelitian ini data akan dibagi menjadi perbandingan 60:40, 70:30 dan 80:20 yang dapat dilihat pada Tabel 3. Metode pembagiannya menggunakan metode *stratified sampling*.

<https://jurnalidpn.badiluhur.ac.id/index.php/Kresna/> 196

3

4

5

Tabel 3. Data

Perbandingan 60:40			
Sentimen	Training	Testing	Total
Positif	182	98	198
Negatif	248	166	414
Total	1430	954	2384
Perbandingan 70:30			
Sentimen	Training	Testing	Total
Positif	1379	591	1970
Negatif	290	124	414
Total	1669	715	2384
Perbandingan 80:20			
Sentimen	Training	Testing	Total
Positif	1576	394	1970
Negatif	331	83	414
Total	1907	477	2384

Training dan Data Testing Pada dataset (dalam menyajikan terhadap ukuran booster (syarat maks))

3.3 Pemodelan

Pada tahapan pemodelan ini, penelitian ini akan dilakukan pengukuran performa klasifikasi dengan dua tahap.

- Pemodelan memakai *split data*. Pada tahapan ini data akan di *split* menjadi 3 rasio yaitu 60:40, 70:30, dan 80:20. Detail dataset dapat dilihat pada Tabel 3.
- Pemodelan menggunakan *cross validation*. Pada tahapan ini data *training* dan data *testing* di *split* secara acak menjadi 10 bagian (*number of folds* = 10).

3.3.1 Hasil Pemodelan Menggunakan Split Data

Pada tabel 4 adalah hasil pemodelan klasifikasi menggunakan *Split Data* dengan melakukan perbandingan antara algoritma KNN, *Naive Bayes* dan SVM. Berdasarkan tampilan pada tabel 4, dapat disimpulkan bahwa nilai akurasi tertinggi ada pada algoritma KNN dengan akurasi sebesar 85.17% dibandingkan dengan algoritma *Naive Bayes* dan SVM.

Tabel 4. Perbandingan Akurasi Algoritma KNN, *Naive Bayes* dan SVM Menggunakan Metode Split Data

Perbandingan	Algoritma		
	KNN	<i>Naive Bayes</i>	SVM
60:40	84.89%	63.37%	82.79%
70:30	85.17%	63.78%	82.52%
80:20	84.91%	62.89%	82.39%

4

5

3.3.2 Hasil Proses Pemodelan Menggunakan Cross Validation

Pada tabel 5 adalah hasil pemodelan klasifikasi menggunakan *Cross Validation* dengan melakukan perbandingan antara algoritma KNN, *Naive Bayes* dan SVM. Berdasarkan tampilan pada tabel 5, dapat disimpulkan bahwa nilai akurasi tertinggi ada pada algoritma KNN dengan akurasi sebesar 84.48% dibandingkan dengan algoritma *Naive Bayes* dan SVM.

Tabel 5. Hasil akurasi Algoritma KNN, *Naive Bayes* dan SVM Menggunakan Metode Cross Validation

Algoritma	Accuracy
K-Nearest Neighbors	84.48%
<i>Naive Bayes</i>	82.75%
Support Vector Machine	83.18%

Berdasarkan Tabel 4, dan Tabel 5, yang berisi hasil pengukuran performa model klasifikasi K-Nearest Neighbors dengan 2 cara yaitu pembagian dataset dengan *split data* dan *cross validation*, maka dapat disimpulkan bahwa pemodelan dengan *split data* 70:30 memiliki akurasi sebesar 85.17% sedangkan dengan pemodelan *cross validation* memiliki akurasi sebesar 84.48%, sehingga akurasi tertinggi yang dihasilkan pada penelitian ini adalah dengan pemodelan dengan *split data* 70:30.

3.3.3 Hasil Perhitungan Algoritma K-Nearest Neighbors

Perhitungan manual dengan algoritma K-Nearest Neighbors ini menggunakan *sample* dari dataset yang diambil secara acak sebanyak 5 data *test* yang dapat dilihat pada Tabel 6, yang dimana *sample* *test* D1 sampai D4 sebagai data *training* dan D5 sebagai data *testing*.

Tabel 6. Contoh *Test*

Sample	Test	Sentimen
D1	Bismillah besok vaksin booster	Positif
D2	Booster jadi syarat Maszid Upiaya Pemerintah Untuk Melindungi Masyarakat	Positif
D3	Bisa kemes sama pusing gitu cek vaksin booster	Negatif
D4	GES TANGANKU SAKIT BGT HABIS VAKSIN BOOSTER HKS GABISA NAK TURUN LENGANNYA SAKIT	Negatif
D5	Apa benar syarat masuk harus udah vaksin booster	?

Setelah menentukan *sample test* yang ingin digunakan, maka tahapan berikutnya adalah menentukan parameter K, kemudian melakukan *preprocessing*, dan menentukan bobot *test* dari 5 *test* menggunakan *Term Weighting* (TF-IDF). Pada contoh Tabel 7, digunakan *k* = 3, karena jumlah data yang digunakan pada Tabel 6, sebanyak 5 data.

Tabel 7. Menentukan Bobot Term

<https://jurnalrpm.budiluhur.ac.id/index.php/Kresna/>

Berikutnya adalah tahap menghitung perkalian skalar. Karena yang akan dicari adalah nilai sentimen dari data *tweet* ke lima (D5) maka hitung perkalian skalar antara data *testing* (D5) dengan data *training* (D1, D2, D3, D4) lalu jumlahkan hasil perkalian skalar pada masing-masing data *training* seperti yang ditunjukkan pada Tabel 8.

	D1	D2	D3	D4
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0.1582	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0.1762	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0.0094	0	0.0084	0.0084	0.0084
0.0094	0.0162	0.0084	0.0084	0.0084

Tabel 9. Panjang Vector Setiap Tweet

<https://jurnal.dipm.budiluhur.ac.id/index.php/Kresna>

3_18_Naskah_Produksi_193-201.pdf 9 / 10 50%

7

8

KRESNA: Jurnal Riset dan Pengabdian Masyarakat
Volume 2, Nomor 2, November, 2022, Hal: 193-201 E-ISSN: 2809-6509

Jika sudah menghitung kemiripan panjang vector dari setiap *tweet*, maka selanjutnya menerapkan persamaan *cosine similarity* sebagai berikut:

$$\text{Cos (D5,D1)} = \frac{0.0094}{0.571 \times 0.993} = -0.0166 \quad (1)$$

$$\text{Cos (D5,D2)} = \frac{0.3168}{0.571 \times 1.507} = -0.3682 \quad (2)$$

$$\text{Cos (D5,D3)} = \frac{0.0094}{0.571 \times 1.851} = -0.0089 \quad (3)$$

$$\text{Cos (D5,D4)} = \frac{0.0094}{0.571 \times 0.994} = -0.0166 \quad (4)$$

Setelah menerapkan persamaan *cosine similarity*, maka selanjutnya adalah mengurutkan berdasarkan jarak dari nilai tertinggi ke yang terendah seperti yang ditunjukkan pada Tabel 10.

Tabel 10. Jarak Berdasarkan Nilai Terendah Ke Tertinggi

D2	D1	D4	D3
0.3682	0.0166	0.0166	0.0089

Jadi dipilih nilai $k=3$ maka akan dipilih 3 nilai *similarity* yang tertinggi, kemudian ambil sebanyak nilai $k (k=3)$ yang paling tinggi tingkat kemiripannya dengan D5 seperti pada Tabel 11.

Tabel 11. *Tweet* Yang Paling Tinggi Kemiripannya

D2	D1	D4
Positive	Positive	Negative

Lalu pilih sentimen yang paling banyak kemunculannya, berdasarkan $k=3$. Sentimen *positive*, diwakili oleh 2 *tweet* yaitu D2 dan D1 sementara sentimen *negative*, diwakili oleh 1 *tweet* yaitu D4. Dari 3 nilai tertinggi tersebut, kelas sentimen yang paling banyak muncul adalah *positive* sehingga *tweet* D5 masuk ke dalam sentimen *positive*.

3.4 Pengujian

Penelitian ini akan melakukan pengujian dengan menggunakan *confusion matrix* yang diperoleh dari tahapan pemodelan dengan menggunakan algoritma KNN dimana pada tabel 4, akurasi dengan *split* data 70:30 lebih tinggi dibandingkan *cross validation*. Sehingga *confusion matrix* yang bisa dilihat pada Tabel 6, adalah hasil dari evaluasi pengujian menggunakan Klasifikasi *K-Nearest Neighbors* dengan dataset yang berjumlah 2384, terdapat 1669 data training, dan 715 data testing.

Tabel 6. *Confusion Matrix*

	True Positive	True Negative	Class Precision
<i>pred positive</i>	584	99	85.11%
<i>pred negative</i>	7	25	78.12%
<i>class recall</i>	98.82%	20.10%	

Perhitungan dari Tabel 6, adalah sebagai berikut:

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{584}{584+99} = 0.8551 = 85.51 \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{584}{584+7} = 0.9882 = 98.82\% \quad (2)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN} = \frac{584+25}{584+25+99+7} = 0.8517 = 85.17\% \quad (3)$$

<https://jurnal.dipn.budihur.ac.id/index.php/Kresna/> 201

3_18_Naskah_Produksi_193-201.pdf 10 / 10 50%

8

9

KRESNA: Jurnal Riset dan Pengabdian Masyarakat
Volume 2, Nomor 2, November, 2022, Hal: 193-201 E-ISSN: 2809-6509

4. KESIMPULAN

Penelitian ini melakukan ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM VAKSIN BOOSTER SEBAGAI SYARAT MUDK 2022 DENGAN ALGORITMA *K-NEAREST NEIGHBORS*. Berdasarkan hasil pengujian dan analisis yang telah dilakukan, menggunakan data dari media sosial *twitter* pada tanggal 27-30 Mei yang sudah dilakukan *preprocessing* dan pemberian label sentimen dengan jumlah dataset yang digunakan 2384 data dengan sentimen positif 1970 data dan sentimen negatif 414 data. Maka dapat disimpulkan bahwa metode algoritma *K-Nearest Neighbors* mendapatkan nilai akurasi sebesar 85.17% saat menggunakan model pengujian *Split Data* dengan perbandingan 70:30 dan saat model pengujian *Cross Validation* dengan $k=9$ mendapatkan nilai akurasi sebesar 84.4%.

DAFTAR PUSTAKA

- [1] F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," *J. Komputasi dan Informatika*, vol. 5, no. 1, pp. 19-25, 2021, doi: 10.31603/komputa.v5i1.5185.
- [2] M. I. Adriansa, R. I. Pratama, K. H. U. Winaha, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengabdian Vaksin COVID-19," *J. Inf. Eng. Educ. Technol.*, vol. 4, no. 2, pp. 90-92, 2020.
- [3] R. Raimanawati, D. Yudhokusuma, P. Hendriarto, Dila Erlanti, and Rini Fitria, "Analisis isi komunikasi publik polisi terkait mudik libur natal dan tahun baru," *J. Komun. Prof.*, vol. 6, no. 1, pp. 26-38, 2022, doi: 10.25139/jkp.v6i1.4403.
- [4] D. Dieti Putri, G. F. Nema, and W. F. Salsitono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terapan*, vol. 10, no. 1, pp. 34-40, 2022, doi: 10.23960/jte.v10i1.2262.
- [5] Z. Nuhila, A. Rahman Ismail, and Z. Alsaba, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/ITSI>
- [6] A. D. Adhi Prita, "Analisis Sentimen pada Ulasan pengguna Aplikasi Ribi Dan Baroksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informatika)*, vol. 8, no. 2, pp. 636-646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [7] D. Astuti, "Penerapan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *J. Informatika, Inf. Sist. Sefor. Eng. Appl.*, vol. 1, no. 2, pp. 60-72, 2019, doi: 10.20895/jinta.v1i2.71.
- [8] E. Wahyu Sholicha, S. Yusra, R. Hamdan, V. Cahya Haridita, T. Rehanza Komputar Jaringana, and P. Tanah Laut, "Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naive Bayes dan K-Nearest Neighbor (Sentiment Analysis of Online Travel Agent Using Naive Bayes and K-Nearest Neighbor)," vol. 3, no. 4, pp. 201-208, 2022.
- [9] A. Apriani, H. Zakkyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF Sistem Pencarian Mahasiswa Baru pada Kampus Swasta," *J. Runggora Inf. Teknol.*, vol. 3, no. 1, pp. 19-27, 2021, doi: 10.30812/rinfo.v3i1.1110.
- [10] S. Khairunnisa, A. Adwijiyo, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidharma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/jmb.v5i2.2815.

<https://jurnal.dipn.budihur.ac.id/index.php/Kresna/> 202