

Sentiment Analysis of Twitter Users' Opinion Towards Face-to-Face Learning*

Analisis Sentimen Tanggapan Masyarakat Pengguna Twitter terhadap
Pembelajaran Tatap Muka

Silmi Annisa Rizki Manaf^{1‡}, Aam Alamudi², dan Anwar Fitrianto³

^{1,2,3}Department of Statistics, IPB University, Indonesia

[‡]Corresponding author: silmiannisarizki@gmail.com

Copyright © 2023 Silmi Annisa Rizki Manaf, Aam Alamudi, and Anwar Fitrianto. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In early 2022, the government allowed face-to-face learning again after approximately one year of online learning. When face-to-face learning will be held again in several areas, the number of Covid-19 has increased and the government has imposed the enforcement of restrictions on community activities. The pros and cons of face-to-face learning also occur on social media, one of them is on Twitter. This study used twitter data for January 30th – February 7th 2022. Opinions on twitter regarding face-to-face learning were studied by sentiment analysis using the binary logistic regression method with sentiment classes being positive and negative. Labeling uses based on the final score of the difference between the number of positive and negative words. The purpose of this study is to determine the public's perception of the policy of implementing face-to-face learning in the era of the Covid-19 on social media especially Twitter. From this study, public's perception tends to be in a negative direction which indicates that they have not agreed enough with the existence of face-to-face learning in the period of February 2022 with the accuracy was 85%, sensitivity was 77%, specificity was 88%, and AUC was 91%.

Keywords: binary logistic regression, face-to-face learning, sentiment analysis, twitter

* Received: Feb 2023; Reviewed: Jul 2023; Published: Oct 2023

1. Pendahuluan

Akhir tahun 2019, tepatnya bulan Desember dunia dihebohkan dengan munculnya suatu virus baru yang menyerang sistem pernapasan, yakni Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). World Health Organization mendeklarasikan bahwa Coronavirus (Covid-19) menjadi Kedaruratan Kesehatan Masyarakat yang Meresahkan Dunia (KKMMD) pada 30 Januari 2020. Virus ini menyebar dari Wuhan, Ibukota Provinsi Hubei China hingga ke seluruh penjuru dunia termasuk Indonesia. Kasus Covid-19 pertama kali ditemukan di Indonesia pada bulan Maret tahun 2020.

Hampir seluruh sektor kehidupan merasakan dampak dari wabah ini tak terkecuali sektor pendidikan. Pemerintah menyampaikan melalui Surat Edaran Menteri Pendidikan dan Kebudayaan Nomor 4 Tahun 2020 yang mengatur tentang Pelaksanaan Kebijakan Pendidikan dalam Masa Darurat Penyebaran Covid-19, seluruh peserta didik diminta untuk mengikuti proses pembelajaran jarak jauh (PJJ) (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, personal communication, 2020). Namun setelah melewati evaluasi selama kurang lebih satu tahun, PJJ dinilai kurang efektif dibandingkan dengan pembelajaran tatap muka (PTM) secara langsung (Statistik, 2021). Berdasarkan hasil evaluasi Kemendikbud, PJJ masih memiliki sejumlah permasalahan seperti halnya tidak memiliki akses teknologi, keterbatasan gawai, jaringan internet, dan media pembelajaran di sejumlah titik.

Pemerintah menyampaikan akan diselenggarakan PTM terbatas dengan tetap menerapkan protokol kesehatan pada pertengahan tahun 2021 dengan melihat situasi dan kondisi yang terjadi. Awal tahun 2022, saat akan diberlangsungkannya kembali PTM, angka Covid-19 kembali meningkat dan membuat pemerintah kembali menerapkan pemberlakuan pembatasan kegiatan masyarakat (PPKM) serta menaikkan status PPKM di sejumlah titik guna menekan lonjakan kasus Covid-19. Hal ini tentunya berpengaruh pada kegiatan PTM yang akan dan sedang berlangsung di beberapa wilayah.

Terjadinya hal tersebut perlu dilakukannya penelitian mengenai implementasi PTM. Bukan hanya dari sisi pendidik dan peserta didik melainkan juga tanggapan dari masyarakat mengenai keberlangsungan PTM. Twitter merupakan media sosial yang cocok digunakan untuk melihat pro dan kontra terhadap kebijakan penerapan PTM di masa pandemi. Topik mengenai PTM sempat menjadi topik yang ramai diperbincangkan karena pernyataan setuju dan tidak setuju dari publik, adanya PPKM, dan isu pemberhentian kembali PTM di beberapa wilayah. Hal ini ramai dibahas tepatnya pada akhir bulan Januari hingga awal bulan Februari tahun 2022 sehingga menjadi latar belakang pemilihan topik mengenai PTM.

Analisis sentimen merupakan salah satu analisis yang cocok digunakan untuk menganalisis sentimen berbasis data twitter yang dapat menjadi satu langkah efisien untuk melihat opini publik terkait kebijakan PTM di Indonesia. Analisis sentimen merupakan cara mengumpulkan pendapat menggunakan jejaring sosial. Di sisi lain, analisis sentimen dapat digunakan untuk mengevaluasi suatu kinerja, pelayanan, dan sebagainya (Syarifuddin, 2020). Penelitian yang dilakukan oleh (Syarifuddin, 2020) membahas tentang opini publik mengenai Covid-19 dengan analisis sentimen dapat menunjukkan dampak positif dan negatif dari Covid-19 sebagai wawasan (insight)

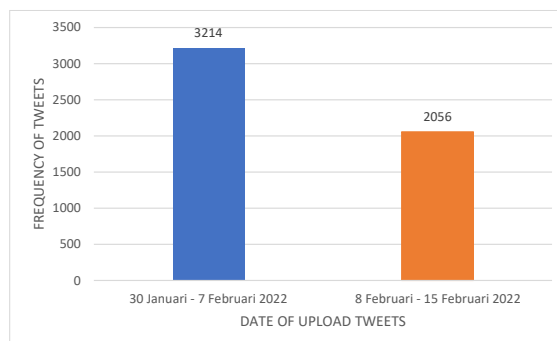
baru untuk masyarakat. Penelitian lainnya melakukan analisis sentimen terhadap PTM pada era pandemi dengan menggunakan algoritma random forest menghasilkan akurasi 76% dan F1-score 68% (Sadi, 2021). Dengan demikian, tujuan dari penelitian ini yaitu untuk mengetahui persepsi masyarakat terhadap kebijakan penerapan pembelajaran tatap muka (PTM) di era pandemi Covid-19 di media sosial.

2. Metodologi

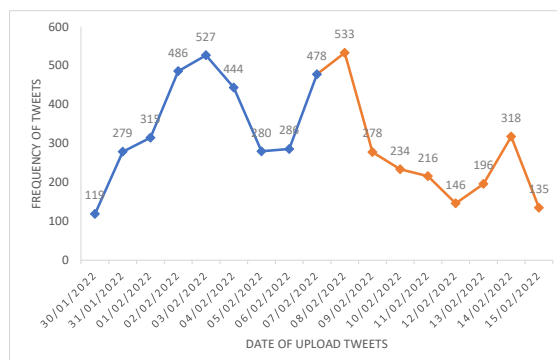
2.1 Data

Data yang digunakan merupakan data hasil penarikan (*data crawling*) dari twitter pada periode 30 Januari hingga 7 Februari 2022 (satu minggu sebelum diberlakukannya kembali PPKM di sejumlah titik, 8 Februari 2022). Menurut definisi secara umum, *crawling* merupakan cara untuk menggali data dari suatu *website* termasuk karakteristik di dalamnya seperti metadata dan *keywords* (Badjrie *et al.*, 2021). Teknik pengambilan data dengan cara ini dibantu oleh *Application Program Interface* (API) sebagai penghubung untuk mengumpulkan data yang ingin digali informasinya. *Crawling* akan menggunakan program yang disebut *crawler* untuk melakukan proses pengambilan data berdasarkan kata kunci yang diinginkan (Saputra, 2017). Data dikumpulkan dengan menggunakan *software* RStudio versi 4.1.1. Batasan dalam proses penarikan data yakni sebagai berikut:

- Kata kunci yang digunakan hanya satu kata kunci yaitu “tatap muka”
- Bahasa yang digunakan adalah twit dengan Bahasa Indonesia
- Unggahan twit yang merupakan *retweet* tidak diikutsertakan untuk menghindari adanya duplikasi data.



Gambar 1: Pemilihan periode penarikan data (*data crawling*)



Gambar 2: Banyaknya twit harian periode 30 Januari - 15 Februari 2022

Gambar 1 menunjukkan perbedaan banyaknya tweet pada dua pekan sebelum dan saat diberlakukannya PPKM. Pekan pertama (30 Januari – 7 Februari 2022), topik mengenai PTM mencapai 3214 tweet dan menurun sebanyak 1158 tweet pada pekan kedua saat PPKM mulai berlangsung (8 Februari – 15 Februari 2022). Gambar 2 menjelaskan bahwa topik mengenai PTM sudah mulai menurun atau tidak mengalami kenaikan tweet yang begitu signifikan. Terlihat pada tanggal 8 Februari ke 9 Februari 2022, banyaknya tweet menurun dan terus menurun di tanggal berikutnya. Hal ini mengindikasikan bahwa periode 30 Januari hingga 7 Februari 2022 tepat dipilih dikarenakan PTM sedang ramai di bahas. Berdasarkan hal tersebut, hasil dari penarikan dengan menggunakan kata kunci “tatap muka” untuk proses analisis selanjutnya data yang digunakan sebanyak 3214 observasi.

Tabel 1 merupakan peubah hasil penarikan data yang digunakan pada tahapan penyiapan data (data preprocessing). Tabel 2 merupakan peubah yang digunakan untuk pemodelan.

Tabel 1: Peubah hasil penarikan data yang digunakan

No	Peubah	Keterangan peubah
1	<i>Text</i>	Isi tweet
2	<i>Created at</i>	Waktu pengunggahan tweet
3	<i>Account created at</i>	Waktu pembuatan akun twitter
4	<i>Followers count</i>	Banyaknya pengikut pada suatu akun
5	<i>Friends count</i>	Banyaknya teman yang diikuti pada suatu akun
6	<i>Statuses count</i>	Banyaknya unggahan tweet yang pada suatu akun

Tabel 2: Peubah yang digunakan pada pemodelan

No	Peubah	Keterangan Peubah
1	Y	Sentimen setiap tweet 1 = Sentimen positif 0 = Sentimen negatif
2	X_k	Bobot kata kunci dari setiap tweet yang diperoleh dari hasil tokenisasi dan pembobotan dengan TF-IDF

2.2 Regresi Logistik

Salah satu metode yang sering digunakan untuk melakukan klasifikasi teks, yaitu regresi logistik. Menurut Pradesa (2019), metode regresi logistik telah banyak digunakan dalam penelitian *text mining* karena memiliki interpretasi probabilistik yang cukup baik. Metode regresi logistik biner akan memodelkan hubungan antara satu atau beberapa peubah prediktor X dengan peubah respon Y berupa biner (Hosmer Jr et al., 2013). Model logistik biner atau disebut sebagai fungsi logit dalam bentuk umum dapat dituliskan sebagai berikut:

$$\pi(x) = P(Y = 1|x) = \frac{e^{g(x)}}{1 + e^{g(x)}} = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} \quad (1)$$

Model regresi logistik biner melibatkan transformasi logit dalam proses pemodelannya. Transformasi logit dilakukan untuk mempertahankan struktur linear dari model di atas yang berbentuk nonlinear. Model linear setelah dilakukannya transformasi logit disebut model regresi logistik sebagai berikut:

$$\text{Logit}(\pi(x)) = g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2)$$

$\pi(x)$ adalah besarnya peluang kelas Y pada observasi ke- i , β_0 adalah intersep regresi logistik, β_j adalah koefisien regresi logistik ke- j , x adalah peubah prediktor, dan p adalah banyaknya peubah prediktor. Pendugaan parameter β_j pada regresi logistik menggunakan metode *Maximum Likelihood Estimation* (MLE) yakni metode pendugaan kemungkinan maksimum dengan nilai penduga parameter β_j didapatkan dari memaksimalkan fungsi kemungkinan (*likelihood function*) (Hosmer Jr *et al.*, 2013).

2.3 Term Frequency-Inverse Document Frequency (TF-IDF)

Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan metode untuk melakukan ekstraksi peubah (*feature extraction*) yang digunakan untuk menentukan bobot pada suatu kata sehingga dapat meningkatkan kinerja dari analisis sentimen. Metode ini akan mentransformasi bentuk teks menjadi bentuk numerik agar dapat digunakan untuk pemodelan. Hasil pembobotan dengan TF-IDF akan diketahui tingkat kepentingan suatu kata terhadap kumpulan twit berdasarkan frekuensi kemunculannya (Sari & Irhamah, 2020). *Term Frequency* (TF) adalah frekuensi munculnya kata dalam suatu dokumen sedangkan *Document Frequency* (DF) adalah banyaknya dokumen yang mengandung kata tertentu. Perhitungan nilai bobot menggunakan rumus TF-IDF sebagai berikut:

$$tf_{t,d} = \frac{n_{t,d}}{\text{jumlah kata dalam dokumen}} \quad (3)$$

$$idf_d = \log\left(\frac{N}{DF_i}\right) \quad (4)$$

$$tfidf_{t,d} = tf_{t,d} \times idf_d \quad (5)$$

$n_{t,d}$ adalah banyaknya kemunculan kata t dalam suatu dokumen d , N adalah banyaknya dokumen secara keseluruhan, DF_i adalah banyaknya dokumen yang mengandung kata t .

2.4 Evaluasi Performa Model Klasifikasi

Evaluasi performa model digunakan untuk mengukur kebaikan suatu model dalam mengklasifikasikan data ke dalam suatu kelas. Matriks konfusi (*confusion matrix*) merupakan matriks berukuran $m \times m$ dengan m adalah banyaknya jumlah kelas klasifikasi yang dapat memperkirakan kemungkinan performa model klasifikasi. Menurut Vujović *et al.*, (2021), titik data dalam matriks konfusi adalah ketika data aktual positif diprediksi positif (*true positive, TP*), data aktual negatif diprediksi positif (*false positive, FP*), data aktual positif diprediksi negatif (*false negative, FN*), dan data aktual negatif diprediksi negatif (*true negative, TN*). Hasil klasifikasi dari setiap kelas yang terklasifikasi benar adalah nilai penjumlahan dari TP dan TN ($TP + TN$) sedangkan hasil klasifikasi palsu adalah dua kemungkinan jenis kesalahan yang dapat dihitung berdasarkan penjumlahan dari FP dan FN ($FP + FN$) (Tyasti *et al.*, 2015).

Tabel 3: Matriks konfusi untuk klasifikasi biner

Kelas prediksi	Kelas aktual	
	Positif	Negatif
Positif	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
Negatif	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Tabel 3 merupakan matriks konfusi dengan dua kelas klasifikasi yakni positif dan negatif. Dari matriks konfusi dapat digunakan untuk menghitung nilai akurasi, sensitivitas, dan Area Under Curve (AUC).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Sensitivitas = \frac{TP}{TP + FN} \quad (7)$$

$$Spesifisitas = \frac{TN}{FP + TN} \quad (8)$$

Penggunaan ukuran kebaikan model dapat digunakan sesuai dengan kebutuhan analisis. Menurut Vujović et al., (2021) nilai terbaik dari akurasi adalah 1 dan yang terburuk adalah 0. Nilai akurasi digunakan untuk mengukur kebaikan model dari keseluruhan pengklasifikasian, sensitivitas cukup efektif untuk mengidentifikasi kelas positif yang terklasifikasi ke dalam kelas positif, spesifisitas cukup efektif untuk mengidentifikasi kelas negatif yang terklasifikasi ke dalam kelas negatif, dan nilai *Area Under Curve* (AUC) merupakan ukuran untuk menentukan kemampuan pengklasifikasian dalam menghindari klasifikasi palsu. AUC dihitung dari luas area di bawah kurva *Receiver Operating Characteristic* (ROC). Ukuran klasifikasi dengan AUC dijelaskan pada Tabel 4 (Ayudhitama & Pujiyanto, 2020).

Tabel 4: Rentang ukuran kualitas klasifikasi

Rentang akurasi	Kualitas klasifikasi
0,90 – 1,00	Sangat baik
0,80 – 0,90	Baik
0,70 – 0,80	Cukup
0,60 – 0,70	Buruk
0,50 – 0,60	Sangat buruk

2.5 Prosedur Analisis Data

1. Pengumpulan data

Proses pengambilan data disebut penarikan data yang didapatkan dari Twitter menggunakan *Application Programming Interface* (API). Dari Twitter API akan diperoleh kunci akses yakni *consumer key*, *consumer secret*, *access token*, dan *access secret* agar bahasa pemrograman yang digunakan dapat terhubung dengan Twitter. Tahapan terakhir yaitu melakukan penarikan data menggunakan RStudio dengan memasukkan kunci akses yang telah didapatkan dan juga kata

kunci yang diinginkan. Data yang sudah selesai ditambang kemudian disimpan dalam format .xlsx untuk selanjutnya dapat dilakukan analisis.

2. Pemilahan data (*data screening*)

Pemilahan data dilakukan dengan menghapus tweet yang berasal dari akun yang dicurigai sebagai akun bot. Akun bot merupakan suatu akun otomatis yang dikontrol oleh program komputer. Beberapa bot menyembunyikan identitas asli dan meniru perilaku kebiasaan manusia untuk memengaruhi pengguna lainnya (Chavoshi & Mueen, 2018). Umumnya untuk melihat antara pengguna asli (*legitimate user*) atau akun bot dapat dilihat dari karakteristik suatu akun tetapi hal ini belum cukup menjamin secara pasti terhadap validitasnya. Penelitian lain menyatakan bahwa tahun pembuatan akun yang sama dapat menandakan umur akun yang baru (Pratama & Rakhmawati, 2019).

$$\text{Umur akun} = \text{Tanggal pembuatan akun} - \text{Tanggal suatu tweet diunggah} \quad (9)$$

Akun yang dicurigai sebagai akun bot dapat dilihat dari banyaknya tweet yang diunggah dalam satu hari. Bila suatu akun melakukan unggahan tweet yang cenderung banyak yakni rata-rata sekitar 100 – 144 tweet per hari maka dapat dicurigai sebagai akun bot.

$$\text{Rata - rata tweet dalam satu hari} = \frac{\text{Statuses count}}{\text{umur akun}} \quad (10)$$

Akun bot memiliki banyaknya teman yang diikuti (*following*) lebih banyak dan lebih acak dibandingkan dengan banyaknya pengikut (*followers*) (Chu *et al.*, 2010). Hal ini dikatakan bahwa akun yang dicurigai sebagai akun bot cenderung akan memperluas jangkauan pertemanan agar mendapatkan tanggapan kembali.

3. Penyiapan data (*data preprocessing*)

Tahapan penyiapan atau praproses data dilakukan untuk membersihkan data agar data siap untuk dianalisis. Tahapan praproses yang dilakukan terdiri dari:

- a. *Case folding* : Mengubah huruf kapital menjadi huruf kecil sehingga data yang digunakan seragam dalam kapitalisasi.
- b. *Text cleans up* : Membersihkan dokumen dari karakter spesial yang tidak dapat terbaca oleh sistem, seperti URL, tautan link, *username*, *emoticon*, angka, *hashtag*, spasi berlebih, dan *e-mail*.
- c. Normalisasi kata : Mengubah kata singkatan dan kata tidak baku menjadi kata baku dari kamus *Colloquial Indonesia Lexicon*.
- d. *Stopwords removal* : Menghilangkan kata yang tidak penting seperti kata sambung dan kata ganti orang.
- e. *Stemming* : Menghilangkan imbuhan pada kata.

4. Penghapusan tweet yang terduplikasi

Tweet yang terduplikasi merupakan tweet yang terkirim lebih dari satu kali sehingga perlu dilakukannya penghapusan tweet dengan tetap mempertahankan tweet pertama.

5. Pelabelan sentimen setiap tweet

Tahapan ini dilakukan dengan memberikan skor sentimen untuk setiap tweet. Apabila skor akhir sentimen > 0 maka dikatakan positif dan skor akhir sentimen < 0 dikatakan negatif. Pelabelan dilakukan berdasarkan rumus yang digunakan pada penelitian Santoso & Nugroho (2019). Adapun rumus yang digunakan sebagai berikut.

$$Skor = \left(\sum \text{kata positif} \right) - \left(\sum \text{kata negatif} \right) \quad (11)$$

6. Tokenisasi dan pembobotan dengan TF-IDF
7. Pemilihan peubah (*feature selection*)
Hasil dari tokenisasi memperoleh peubah prediktor yang banyak. Ukuran peubah yang besar ini dapat memperlambat proses komputasi terlebih lagi tidak semua peubah yang dihasilkan relevan. Sehingga dilakukannya seleksi peubah untuk mencari peubah yang memiliki keterkaitan yang cukup tinggi dengan menggunakan *information value*. *Information value* merupakan seleksi fitur yang cukup optimal untuk model klasifikasi dengan regresi logistik.
8. Pemodelan menggunakan regresi logistik biner
Pemodelan dilakukan menggunakan regresi logistik biner dengan peubah respon Y merupakan kelas sentimen dari setiap tweet yang diberi kode sebagai berikut:

$$Sentimen (Y) = \begin{cases} 1 = Sentimen positif \\ 0 = Sentimen negatif \end{cases}$$

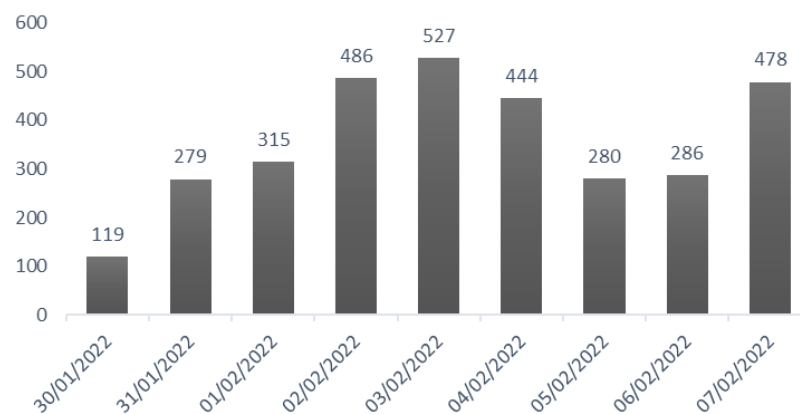
Pemodelan menggunakan *k-folds cross validation* dengan $k = 10$. Validasi silang dengan 10 lipatan (*folds*) merupakan hal yang umum digunakan dalam penambangan data serta pemilihan $k = 10$ merupakan metode pemilihan model terbaik. Penggunaan $k = 10$ lebih efisien dari sisi komputasi dan secara efektif dapat meningkatkan ketepatan estimasi performa model dengan tetap mempertahankan bias yang kecil (Kuhn et al., 2013).

9. Pengevaluasian performa model berdasarkan nilai akurasi, sensitifitas, spesifisitas, dan *Area Under Curve* (AUC)
Hasil dari pemodelan akan didapatkan ukuran performa model untuk setiap lipatan (*folds*). Hasil performa model akhir secara keseluruhan didapatkan dari perhitungan rata-rata dari setiap nilai akurasi, sensitivitas, spesifisitas, dan AUC.
10. Pembuatan awan kata (*word cloud*)
Visualisasi awan kata (*word cloud*) merupakan cara untuk dapat menggambarkan kata yang paling sering muncul dalam suatu tweet yang di analisis. Semakin besar ukuran huruf atau kata yang ditampilkan artinya kata tersebut semakin sering muncul dalam data. Visualisasi ini tepat digunakan untuk cepat menemukan kata yang sering muncul meskipun tidak dapat menunjukkan frekuensi kata-kata yang muncul dalam suatu teks yang dianalisis (Adiyana & Hakim, 2015).
11. Penarikan simpulan dan saran berdasarkan hasil pemodelan

3. Hasil dan Pembahasan

3.1 Data dan Praproses Data

Data hasil penambangan didapatkan 3214 observasi yang telah sesuai dengan kriteria batasan yang telah ditentukan. Dalam penelitian ini digunakan satu kata kunci karena kata kunci tersebut sudah cukup menggambarkan topik secara keseluruhan.

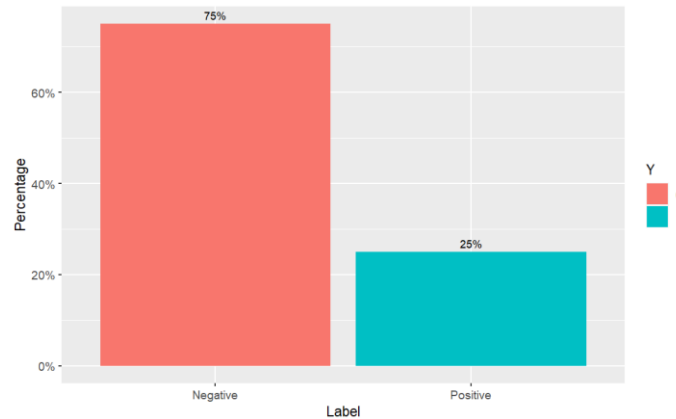


Gambar 3: Frekuensi tweet per hari

Gambar 3 menunjukkan grafik frekuensi tweet per hari dimulai dari tanggal 30 Januari 2022 sampai 7 Februari 2022. Berdasarkan grafik tersebut, tweet paling banyak diunggah mengenai pembelajaran tatap muka pada tanggal 3 Februari 2022 sebanyak 527 tweet dalam satu hari. Berdasarkan hasil analisis data, tanggal 3 Februari 2022 pemerintah mulai meluncurkan surat edaran mengenai PTM setelah naiknya kasus Covid-19 di awal tahun. Pemerintah menetapkan PTM menjadi 50% dan beberapa pemerintah setempat lainnya ada yang memberhentikan PTM 100%. Data observasi akhir setelah melalui tahapan penghapusan akun yang dicurigai sebagai akun bot, praproses data, dan penghapusan data yang terduplikasi menjadi 2936 observasi.

3.2 Pelabelan

Pelabelan dilakukan terhadap 2936 observasi menggunakan Persamaan 11. Dari hasil pemberian skor dihasilkan sebanyak 423 data bersentimen positif dan 1272 data bersentimen negatif. Sebanyak 1241 data bersentimen netral. Kelas netral tidak diikutsertakan karena mengacu pada beberapa peneliti sebelumnya, salah satunya yang menyatakan bahwa dari beberapa survei literatur, sebagian besar peneliti fokus pada dua kelas yaitu positif dan negatif. Hal lain disebutkan bahwa metode konvensional menunjukkan akurasi prediksi yang rendah untuk masalah 3 kelas yang mengandung kelas netral dikarenakan adanya ketidakseimbangan data yang disebabkan oleh jumlah data kelas netral (Ray et al., 2021). Sehingga banyaknya observasi yang digunakan dalam penelitian ini menjadi 1695 observasi. Data bersentimen netral tidak diikutsertakan dalam penelitian karena dianggap cenderung tidak dapat terlihat arah opini yang diinginkan baik ke arah opini positif atau negatif. Tidak ada makna khusus untuk digunakan dalam penelitian.



Gambar 4: Bar chart distribusi kelas pelabelan

Gambar 4 merupakan ringkasan dari pelabelan yang telah dilakukan. Secara eksplorasi melalui bar chart didapatkan bahwa tweet mengenai PTM cenderung ke arah sentimen negatif yang lebih besar dibandingkan dengan sentimen positif. Antar kelas positif dan negatif memiliki rasio 1:3. Sehingga berdasarkan penelitian yang dilakukan oleh Krawczyk (2016), ketidakseimbangan data berada pada rasio mulai dari 1:4 hingga 1:100 maka data pada penelitian ini masih dapat dikatakan seimbang dan tidak dilakukannya penanganan ketidakseimbangan data.

3.3 Tokenisasi dan Pembobotan TF-IDF

Tokenisasi dilakukan untuk proses pemisahan suatu kalimat menjadi potongan kata. Kalimat dipecah menjadi bentuk token (kata) menggunakan pembatas atau delimiter spasi. Teknik ini disebut dengan tokenisasi unigram. Setiap potongan kata yang dihasilkan dari proses tokenisasi dilanjutkan dengan pemberian bobot menggunakan TF-IDF. Pembobotan dengan TF-IDF dilakukan untuk mentransformasi bentuk teks menjadi bentuk peubah numerik agar dapat dilakukan pemodelan.

Tabel 5: Contoh tweet hasil tokenisasi

Kalimat	Token	Indeks
moga tatap muka	“moga”, “tatap”, “muka”	[1, 2, 3]
besok pagi tatap muka	“besok”, “pagi”, “tatap”, “muka”	[4, 5, 2, 3]

Tabel 6: Contoh pemberian bobot dengan TF-IDF

Tweet ke-	Kata							Sentimen
	aaa	...	ajar	...	pandemi	...	zzzzzz	
1	0	...	0	...	0	...	0	Negatif
2	0	...	0,0865	...	0	...	0	Positif

Tabel 5 menunjukkan bahwa suatu kalimat yang memiliki kata yang sama akan masuk ke dalam indeks yang sama. Kata “tatap” pada tweet kedua masuk ke dalam indeks 2 dan kata “muka” masuk ke dalam indeks 3 sedangkan kata “besok” dan “pagi” masuk ke dalam indeks baru (indeks 4 dan 5) karena belum ada kedua kata tersebut pada tweet sebelumnya. Dari hasil tokenisasi didapatkan sebanyak 4589 kata.

Tabel 6 menunjukkan kata-kata yang terbentuk akan diberi bobot menggunakan metode TF-IDF. Twit pertama tidak mengandung kata “aaa”, “ajar”, “pandemi”, dan “zzzz” sehingga setiap kata pada twit pertama memiliki bobot nol. Sedangkan pada twit kedua mengandung kata “ajar” sehingga kata “ajar” pada twit kedua mengandung bobot 0,0865. Hasil pembobotan dengan TF-IDF merupakan hasil perkalian antara TF dan IDF sehingga memungkinkan bernilai nol apabila TF maupun IDF bernilai nol atau dikarenakan twit tersebut tidak mengandung kata ke-t. Hasil dari pembobotan akan menghasilkan peubah prediktor X yang digunakan untuk melakukan pemodelan.

3.4 Pemilihan Peubah (Feature Selection)

Banyaknya peubah setelah melalui tahap tokenisasi yaitu sebanyak 4589 peubah. Hasil dari tokenisasi dan pembobotan dengan TF-IDF ini akan menjadi peubah X yang digunakan untuk pemodelan. Ukuran data besar yaitu 2936 baris \times 4589 kolom dapat memperlambat proses komputasi terlebih lagi tidak semua peubah yang dihasilkan relevan terhadap kata. Sehingga akan dilakukannya seleksi fitur untuk mengurangi peubah prediktor X .

Tabel 7: *Information value* peubah prediktor

No	Peubah	<i>Information value</i>
1	muka	0,3359
2	tatap	0,3159
...
4589	prasetyani	0,0003

Information value $< 0,02$ mengindikasikan peubah prediktor tidak cukup berguna dalam pemodelan (Siddiqi, 2012). Secara definisi, *information value* (IV) merupakan salah satu metode seleksi fitur yang dapat digunakan untuk memilih peubah yang relevan. Perhitungannya berkaitan erat dengan *weight of evidence* (WoE). WoE dipandang sebagai suatu ukuran untuk melihat kategori dari peubah prediktor X yang cenderung berasosiasi dengan kelas dari peubah respon Y (Sartono et al., 2021). Nilai IV dapat dihitung melalui persamaan berikut ini.

$$IV = \sum_{i=1}^n (Distribusi\ positif - Distribusi\ negatif) \times \ln \left(\frac{Distribusi\ positif}{Distribusi\ negatif} \right) \quad (12)$$

Hasil perhitungan IV menunjukkan nilainya selalu bernilai positif yang mengindikasikan jika nilainya semakin besar pada suatu peubah artinya peubah tersebut semakin dapat membedakan antara kelas positif dan negatif (Sartono et al., 2021). Sehingga dari 4589 peubah prediktor yang didapatkan, hanya 278 yang memiliki *information value* $> 0,02$. Berikut gambaran data yang akan digunakan dalam pemodelan, terdapat peubah respon Y yang terdiri atas kelas positif dan negatif dan 278 peubah prediktor X .

Tabel 8: Peubah yang digunakan untuk pemodelan

Twit ke- i	Y	muka	tatap	pembelajaran	...	tuju
1	0	0,0792	0,0786	0,1682	...	0
2	0	0,1218	0,1209	0	...	0
...
1695	0	0,0320	0,0317	0,0679	...	0

3.5 Referensi Pelabelan Lain

Peneliti telah melakukan pemodelan dengan menggunakan beberapa metode pelabelan yang berbeda untuk mengkaji analisis sentimen sehingga dapat dijadikan sebagai referensi untuk peneliti lain. Pelabelan pertama dilakukan secara manual. Kelas positif merupakan kelas yang mengandung pernyataan setuju dan dukungan sedangkan kelas negatif yakni kelas yang mengandung pernyataan yang tidak setuju, ejekan, dan kontra terhadap topik pembicaraan. Metode pelabelan ini dilakukan berdasarkan subjektivitas peneliti. Sehingga sangat memungkinkan tingkat kesalahan manusia (*human error*) yang tinggi. Pelabelan kedua disebut pelabelan formula 1 dengan menggunakan Persamaan 11. Pelabelan ketiga disebut pelabelan formula 2 didapatkan dari perhitungan skor sentimen berdasarkan kamus sentimen, boosterwords, dan kamus negasi. Perhitungan skor sentimen dilakukan dengan menjumlahkan skor pada kata-kata yang terdapat di dalam twit berdasarkan kata yang terkandung di dalam ketiga kamus tersebut (Aliyah et al., 2020). Pelabelan keempat dilakukan dengan memvalidasi ketiga hasil pelabelan lalu mengategorikan kelas positif dan negatif. Dari tiga metode pelabelan akan dihasilkan satu label akhir yang selanjutnya akan digunakan sebagai pelabelan validasi. Keempat metode pelabelan tersebut selanjutnya dilakukan pemodelan dengan regresi logistik biner dan dihasilkan nilai rata-rata dari sepuluh lipatan (*folds*) ukuran performa model untuk setiap metode pelabelan. Hasilnya sebagai berikut.

Tabel 9: Evaluasi performa model setiap metode pelabelan

Metode pelabelan	Akurasi	Sensitivitas	Spesifisitas	AUC
Pelabelan manual	0,5877	0,3404	0,8717	0,6846
Pelabelan formula 1	0,8519	0,7677	0,8797	0,9125
Pelabelan formula 2	0,6899	0,7164	0,6456	0,7837
Pelabelan validasi	0,5981	0,2586	0,9069	0,7093

Hasil performa model setiap metode pelabelan, setiap model memberikan nilai performa model yang cukup. Sehingga dengan kata lain, untuk mengkaji analisis sentimen dengan menggunakan salah satu ataupun keempat metode pelabelan ini sudah cukup memberikan nilai AUC yang cukup baik yakni di atas 0,70. Tetapi dengan menggunakan metode pelabelan formula 1 memberikan pemodelan yang paling baik di antara model terbaik lainnya (di atas 0,75 untuk setiap nilai performa model).

3.6 Pemodelan dan Evaluasi Performa Model

Proses pemodelan menggunakan k-folds cross validation dengan $k = 10$. Dari 1695 data observasi karena akan dibagi ke dalam 10 folds maka terdapat sekitar 1526 yang menjadi data latih (90%) dan 169 yang menjadi data uji (10%) dengan iterasi sebanyak 10 kali. Data penelitian di bagi ke dalam sepuluh bagian. Model pertama adalah data bagian pertama sebagai data uji sedangkan sembilan bagian data lainnya sebagai data latih, dan seterusnya sampai model kesepuluh. Data latih digunakan untuk proses membangun model selanjutnya akan dievaluasi dengan data uji.

Tabel 10: Performa model setiap lipatan (fold)

<i>Fold ke-i</i>	Akurasi	Sensitivitas	Spesifisitas	AUC	TP	FP	TN	FN
1	0,8353	0,6341	0,8992	0,8966	26	13	116	15
2	0,8588	0,8500	0,8615	0,9483	34	18	112	6
3	0,8059	0,7429	0,8222	0,8870	26	24	111	9
4	0,8647	0,6905	0,9219	0,9122	29	10	118	13
5	0,9118	0,8750	0,9262	0,9575	42	9	113	6
6	0,8580	0,8163	0,8750	0,9034	40	15	105	9
7	0,8935	0,7872	0,9344	0,9435	37	8	114	10
8	0,8284	0,7750	0,8450	0,8899	31	20	109	9
9	0,8343	0,7500	0,8571	0,8835	27	19	114	9
10	0,8284	0,7556	0,8548	0,9032	34	18	106	11

Hasil performa model akhir secara keseluruhan akan dihitung berdasarkan perhitungan rata-rata dari setiap nilai akurasi, sensitivitas, spesifisitas, dan AUC dari kesepuluh *folds*.

Tabel 11: Evaluasi klasifikasi berdasarkan data latih dan data uji

Data	Akurasi	Sensitivitas	Spesifisitas	AUC
Data latih	0,9655	0,9185	0,9810	0,9935
Data uji	0,8519	0,7677	0,8797	0,9125

Tabel 10 menunjukkan bahwa klasifikasi teks menggunakan metode regresi logistik biner memiliki nilai akurasi, sensitivitas, spesifisitas, dan AUC yang cukup tinggi (di atas 0,70). Data *tweet* yang memuat kata kunci "tatap muka" tidak mengalami gejala *overfitting*. Hal ini ditunjukkan pada Tabel 11, nilai akurasi dari data latih dan data uji tidak jauh berbeda serta nilai akurasi pada data latih yang lebih besar dibandingkan dengan data uji. Hal ini dapat disebabkan karena banyaknya data latih yang digunakan menghasilkan model yang terbentuk dapat menangani lebih banyak keberagaman yang ada pada data sehingga mampu mengklasifikasikan dengan lebih baik ketika dilakukan pengujian pada data uji.

Hasil pengklasifikasian kelas sentimen dengan regresi logistik biner menunjukkan bahwa model sudah cukup baik dalam mengklasifikasikan kelas sentimen *tweet* untuk setiap ukuran performa model. Evaluasi performa model akhir dapat dilihat berdasarkan evaluasi klasifikasi berdasarkan data uji untuk metode pelabelan formula 1. Nilai rata-rata akurasi, sensitivitas, spesifisitas, dan AUC dari 10 *folds* didapatkan nilai akurasi 85,19%, sensitivitas 76,77%, spesifisitas 87,97%, dan AUC 91,25%. Nilai akurasi dapat diinterpretasikan bahwa model menghasilkan klasifikasi kelas sentimen 85,19% benar sesuai kelas sentimen aslinya.

Tabel 12: Matriks konfusi akhir secara keseluruhan

Kelas prediksi	Kelas aktual	
	Positif	Negatif
Positif	326	154
Negatif	97	1118

Tabel 12 menunjukkan matriks konfusi akhir secara keseluruhan yakni penjumlahan dari kesepuluh nilai TP, FP, TN, dan FN dari hasil pemodelan dengan regresi logistik biner menggunakan metode *10-folds cross validation* dengan metode pelabelan formula 1. Matriks konfusi menunjukkan bahwa kemungkinan kesalahan pengklasifikasian kelas masih tergolong cukup rendah yakni 251 atau sekitar 14,81% (97 terklasifikasi ke dalam nilai FN dan 154 terklasifikasi ke dalam nilai FP) jika dibandingkan dengan banyaknya tweet yang terklasifikasi sesuai dengan kelas data aktualnya yakni 1444 atau sekitar 85,19% (326 terklasifikasi ke dalam nilai TP dan 1118 terklasifikasi ke dalam nilai TN).

Tabel 13: Contoh data aktual dan hasil klasifikasi

Teks	Data aktual	Hasil klasifikasi
seru kadang tangis tugas banget tugas ya seru banget asik mapel bikin kaget kalo tatap muka gampang	Positif	Negatif
pjj susah fokus ganggu orang rumah kadang suruh pas ajar efektif tatap muka guru ngejelasinnya mudah paham sosialisasi sekolah	Negatif	Negatif

Tabel 13 merupakan contoh hasil pengklasifikasian, tweet pertama bersentimen positif namun model mengklasifikasikan ke dalam sentimen negatif. Model belum dapat secara tepat mengklasifikasikan tweet ke dalam kelas sentimen yang sesuai karena masih terdapat beberapa kata-kata yang cenderung negatif pada tweet pertama tetapi sebenarnya tweet pertama menunjukkan pernyataan setuju terhadap PTM. Contoh pada tweet kedua menunjukkan kesesuaian terhadap hasil pengklasifikasian, yakni tweet yang bersentimen negatif dan terklasifikasi ke dalam sentimen negatif.

3.7 Pembentukan Awan Kata



Gambar 5: Word cloud dari tweet yang berkaitan dengan PTM (a) positif, (b) negatif, dan (c) keseluruhan

Visualisasi awan kata (*word cloud*) merupakan suatu cara sederhana untuk menggambarkan kata yang paling sering muncul di dalam kumpulan tweet atau yang paling sering di bahas oleh pengguna twitter mengenai PTM. Semakin besar ukuran huruf artinya kata tersebut semakin sering muncul dan sebaliknya. Gambar 5 (a)

menunjukkan visualisasi kata pada sentimen positif. Berdasarkan hasil di atas, kata yang paling sering muncul adalah “kuliah”, “bijak”, dan “evaluasi”. Gambar 5 (b) menunjukkan visualisasi kata pada sentimen negatif. Berdasarkan hasil visualisasi didapatkan kata-kata yang sering muncul yakni “online”, “daring”, dan “omicron”. Dari hasil visualisasi Gambar 5 (c) ditemukan topik yang hangat di bahas secara keseluruhan yakni “pembelajaran”, “tatap”, “muka”, “sekolah”, “ptm”, “covid”. Kata-kata tersebut memiliki jumlah kemunculan yang tertinggi di setiap tweet.

Hasil visualisasi ini menunjukkan pernyataan mendukung dan tidak mendukung dengan adanya PTM memiliki alasan tersendiri. Sentimen negatif lebih besar daripada sentimen positif menunjukkan bahwa masyarakat belum sepenuhnya setuju dengan adanya PTM terlebih diakibatkan karena harus melakukan adaptasi kembali dengan adanya sekolah tatap muka dan masyarakat cenderung menganggap dengan diberlakukannya kembali PTM menjadi wadah sebagai menyebarnya Covid-19 kembali. Sedangkan masyarakat yang setuju dengan adanya PTM beranggapan bahwa PTM menjadikan suatu mata pelajaran lebih mudah dimengerti dan siap untuk bersosialisasi kembali. Meskipun demikian, dari hasil visualisasi ini, *word cloud* sentimen positif menunjukkan kata “evaluasi” cukup memiliki ukuran yang besar. Sehingga dengan diberlakukannya kembali PTM diharapkan dapat dievaluasi lebih lanjut terkait sistem yang sedang berjalan agar orang tua, pendidik, tenaga pendidik, dan masyarakat sekitar merasa lebih aman dan nyaman dengan adanya sekolah *offline*.

4. Simpulan dan Saran

Persepsi masyarakat berdasarkan komentar di twitter mengenai PTM menunjukkan bahwa masyarakat belum sepenuhnya siap dengan adanya PTM di era pandemi saat ini khususnya pada saat periode bulan Februari lalu ketika angka Covid-19 kembali meningkat. Analisis sentimen dilakukan untuk mempertemukan opini dari masyarakat. Dari opini ini menunjukkan bahwa masyarakat baik murid, orang tua murid, dan pengguna twitter dari berbagai kalangan sudah cukup beradaptasi dengan adanya PJJ, sehingga jika harus dilakukannya kembali PTM maka harus melakukan adaptasi kembali. Dari hasil ini diharapkan adanya kebijakan atau inovasi baru dalam masa transisi menuju PTM agar kegiatan ini tetap berjalan dengan aman dan nyaman.

Hasil klasifikasi teks menggunakan regresi logistik biner untuk mengkaji analisis sentimen cukup memberikan ukuran performa model yang cukup tinggi yakni nilai akurasi 85,19%, sensitivitas 76,77%, spesifisitas 87,97%, dan AUC 91,25%. Matriks konfusi akhir secara keseluruhan menunjukkan bahwa kemungkinan kesalahan pengklasifikasian kelas yang ditunjukkan dari nilai *false positive* dan *false negative* masih tergolong cukup rendah jika dibandingkan dengan banyaknya tweet yang terklasifikasi sesuai dengan kelas data aktualnya (ke dalam kelasnya) yakni ditunjukkan dari nilai *true positive* dan *true positive*. Selain itu juga ketepatan akurasi dari klasifikasi sentimen masyarakat menggunakan regresi logistik biner terhadap PTM di era pandemi dapat dilihat dari nilai sensitivitas dan spesifisitas yang memberikan nilai cukup tinggi. Sehingga model regresi logistik biner memberikan ketepatan klasifikasi yang cukup baik dalam mengkaji analisis sentimen.

Saran yang dapat diberikan kepada peneliti selanjutnya mengenai analisis sentimen pada twitter yakni dapat lebih mengoptimalkan dalam praproses data, seperti halnya membersihkan akun yang benar merupakan akun bot, menambah kata di dalam daftar *stopwords*, serta menambah daftar kata tidak baku untuk menangani

masalah kesalahan penulisan pada tahap normalisasi kata agar hasil praproses data menjadi lebih bersih tanpa harus melakukan pengecekan kembali secara manual. Lainnya dapat menggunakan metode pelabelan seperti yang telah dilakukan pada penelitian. Terakhir dapat menggunakan metode klasifikasi teks lain seperti *Support Vector Machine*, *Naïve Bayes*, *Random Forest* agar dapat membandingkan kinerja dari algoritma tersebut.

Daftar Pustaka

- Adiyana, I., & Hakim, R. (2015). Implementasi text mining pada mesin pencarian Twitter untuk menganalisis topik-topik terkait “KPK dan Jokowi”. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS*, 570–581.
- Aliyah, S. F., Yasin, H., Suparti, B. W., & Widiharis, T. (2020). Analisis sentimen PT TIKI Jalur Nugraha Ekakurir (PT TIKI JNE) pada media sosial Twitter menggunakan model feed forward neural network. *Statistika*, 8(2): 103–113.
- Ayudhitama, A. P., & Pujianto, U. (2020). Analisa 4 algoritma dalam klasifikasi liver menggunakan rapidminer. *Jurnal Informatika Polinema (JIP)*, 6(2): 1–9.
- Badjrie, S. H., Pratiwi, O. N., & Anggana, H. D. (2021). Analisis sentimen review customer terhadap produk indihome dan first media menggunakan algoritma convolutional neural network. *eProceedings of Engineering*, 8(5): 9049–9061.
- Chavoshi, N., & Mueen, A. (2018). Model bots, not humans on social media. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 178–185.
- Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: Human, bot, or cyborg?. *Proceedings of the 26th Annual Computer Security Applications Conference*, 21–30.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression. 3rd edition*. New Jersey (NJ): John Wiley & Sons.
- Kementrian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2020). *Surat Edaran Mendikbud Nomor 4 Tahun 2020 Tentang Pelaksanaan Kebijakan Pendidikan dalam Masa Darurat Penyebaran Corona Virus Disease (Covid-19)* [Personal communication].
- Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4): 221–232.
- Kuhn, M., Johnson, K., & others. (2013). *Applied Predictive Modeling*. New York (USA): Springer.
- Pradesa, P. S. (2019). *Analisis sentimen pada twitter tentang kepuasan pelanggan indihome dengan metode logistic regression*.
- Pratama, P. G., & Rakhmawati, N. A. (2019). Social bot detection on 2019 Indonesia president candidate’s supporter’s tweets. *Procedia Computer Science*, 161: 813–820.
- Ray, B., Garain, A., & Sarkar, R. (2021). An ensemble-based hotel recommender

- system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing*, 98: 106935.
- Sadi, N. R. (2021). *Analisis sentimen terhadap pembelajaran tatap muka pada era pandemi covid-19 menggunakan algoritma random forest* [skripsi]. Bandung (ID): UIN Sunan Gunung Djati Bandung.
- Santoso, E. B., & Nugroho, A. (2019). Analisis sentimen calon presiden Indonesia 2019 berdasarkan komentar publik di Facebook. *Jurnal Eksplora Informatika*, 9(1): 60–69.
- Saputra, P. Y. (2017). Implementasi teknik crawling untuk pengumpulan data dari media sosial Twitter. *Dinamika Dotcom: Jurnal Pengembangan Manajemen Informatika dan Komputer*, 8: 160–168.
- Sari, E. D. N., & Irhamah, I. (2020). Analisis sentimen nasabah pada layanan perbankan menggunakan metode regresi logistik biner, Naïve Bayes classifier (NBC), dan support vector machine (SVM). *Jurnal Sains Dan Seni ITS*, 8(2): D177–D184.
- Sartono, B. & others. (2021). *Teknik Eksplorasi Data yang Harus dikuasai Data Scientist*. Bogor (ID): PT Penerbit IPB Press.
- Siddiqi, N. (2012). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring* (Vol. 3). New Jersey (NJ): John Wiley & Sons.
- Statistik, B. P. (2021). *Statistik Pendidikan 2021*. Jakarta (ID): Badan Pusat Statistik.
- Syarifuddin, M. (2020). Analisis sentimen opini publik mengenai covid-19 pada Twitter menggunakan metode Naïve Bayes dan KNN. *Inti Nusa Mandiri*, 15(1): 23–28. <https://doi.org/10.33480/inti.v15i1.1347>
- Tyasti, A. E., Ispriyanti, D., & Hoyyi, A. (2015). Algorithm iterative dichotomiser 3 (id3) untuk mengidentifikasi data rekam medis (studi kasus penyakit diabetes mellitus di Balai Kesehatan Kementrian Perindustrian, Jakarta). *Jurnal Gaussian*, 4(2): 237–246. <https://doi.org/10.14710/j.gauss.v4i2.8422>
- Vujović, Ž. & others. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, 12(6): 599–606.