

ANALISIS SENTIMEN PADA *TWITTER* TERKAIT *NEW NORMAL* DENGAN METODE *NAÏVE BAYES CLASSIFIER*

Imelya Susianti¹⁾, Sri Soerya Ningsih²⁾, M. Al Haris³⁾, Tiani Wahyu Utami⁴⁾

¹⁾Mahasiswa Universitas Muhamadiyah Semarang
email: imelyasusianti@gmail.com

²⁾Mahasiswa Universitas Muhamadiyah Semarang
email: srisurya53.ss@gmail.com

³⁾Dosen Universitas Muhamadiyah Semarang
email: alharis3@gmail.com

⁴⁾Dosen Universitas Muhamadiyah Semarang
email: tianiunimus@gmail.com

Abstract

The rapid development of information technology has made it easy for the public to access information quickly, one of which is on social media. Social media, especially Twitter, is now a very popular communication tool among internet users. With social media twitter, there is information data that can be processed into sentiment analysis. This study aims to classify positive and negative public sentiments on twitter related to the new normal during the pandemic. The method used in this research is the Naïve Bayes Classifier and the output of this system is visualization in the form of wordcloud. The data used were 3000 tweets consisting of 2409 positive tweets and 591 negative tweets. Sentiment analysis using the Naïve Bayes Classifier with TF-IDF weighting produces an average accuracy value of 87.33%.

Keywords: *New Normal, Naïve Bayes Classifier*

1. PENDAHULUAN

Tahun 2020 dunia digemparkan dengan adanya suatu wabah virus berbahaya yang menyerang manusia yaitu virus corona. Virus ini pertama kali muncul di Wuhan Cina dan menyebar hingga ke negara lainnya termasuk negara Indonesia. Berbagai upaya telah dilakukan di Indonesia demi mengantisipasi virus tersebut dari *social distancing*, *lockdown* hingga *New Normal*. Banyak sekali respon atau tanggapan masyarakat terkait *New Normal*. Khususnya di media sosial *Twitter*.

Berdasarkan data dari Drone Emprit pada 2020, dalam tiga hari terakhir setelah pemerintah mengumumkan tentang *New Normal*, terjadi lonjakan percakapan dengan kata *New Normal* pada media sosial *Twitter* sehingga memunculkan berbagai komentar. Berbagai komentar yang bermunculan ada yang berupa opini positif ada pula opini negatif. Maka untuk mengetahui opini-opini dari komentar yang ada di *Twitter* harus diolah dengan mengklasifikasikan menjadi opini positif atau opini negatif. Salah satu analisis yang dapat digunakan yaitu analisis sentimen. Analisis sentimen digunakan untuk mendapatkan sebuah uraian umum tentang kualitas sebuah layanan, apakah layanan tersebut cenderung mendapatkan nilai positif, negative maupun netral (Istiqomah, Maharani & Adiwijaya, 2014).

Penelitian dengan menggunakan analisis sentimen telah banyak dilakukan, beberapa diantaranya yaitu yang dilakukan oleh Alfredo Gormantara (2020) dengan judul : "Analisis Sentimen Terhadap *New Nomal Era* di Indonesia pada *Twitter* Menggunakan Metode *Support Vector Mechine*" dengan hasil akurasi negatif atau positif sebesar 76,5%, *recall* 90,91% dan *precesion* 70,80%. Penelitian yang dilakukan oleh Syakuro, Abdan (2017) dengan judul : "Analisis Sentimen Masyarakat terhadap *E-Commerce* pada Media

Sosial menggunakan Metode *Naïve Bayes Classifier (NBC)* dengan Seleksi Fitur *Information Gain(IG)*” mendapatkan hasil tingkat akurasi sebesar 88,8%. Serta penelitian yang dilakukan oleh Servasius Dwi Harijiatno (2019) dengan judul : “Analisis Sentimen pada Twitter menggunakan *Multinomial Naïve Bayes*” hasil yang diperoleh akurasi sebesar 72,941% tanpa menggunakan *k-Fold Cross Validation* dan dengan menggunakan *k-Fold Cross Validation* sebesar 71,601%, 70,72% dan 70,68%.

Dalam penelitian ini, metode pengklasifikasian yang digunakan yaitu *Naïve Bayes Classifier* karena metode ini merupakan metode dengan teknologi *pre-processing* dalam klasifikasi fitur, yang dapat menambah skalabilitas, akurasi dan efisiensi dalam proses klasifikasi sebuah teks serta memiliki kelebihan yaitu algoritmanya sederhana tetapi memiliki akurasi yang tinggi (Rish, 2006).

2. KAJIAN LITERATUR

New Normal

New Normal adalah perubahan perilaku untuk tetap menjalankan aktivitas normal, namun dengan menerapkan protokol kesehatan dengan tujuan mencegah terjadinya penularan Covid-19 (Yuli Budirahayu, 2020) dengan mengadopsi perilaku hidup berbeda agar menekan risiko penularan virus.

Twitter

Twitter adalah suatu situs *web* yang merupakan layanan dari *microblog*, yaitu suatu bentuk *blog* yang membatasi ukuran setiap posnya, yang memberikan fasilitas bagi pengguna untuk dapat menuliskan pesan dalam *twitter update* hanya berisi 140 karakter.

Text Mining

Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen.

Text Preprocessing

Preprocessing merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Secara umum beberapa proses yang dilakukan pada tahap *preprocessing* adalah sebagai berikut :

a *Case Folding*

Case folding adalah proses mengubah semua huruf besar atau kapital dalam *Twitter* menjadi huruf kecil.

b *Tokenizing*

Tokenizing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini memisahkan setiap kata yang menyusun suatu dokumen.

c *Filtering*

Filtering adalah proses mengambil kata-kata penting dari hasil token. Pada proses ini akan dilakukan pembersihan *tweet* dari spesial karakter, *URL link*, *username*, serta *emoticon*. Proses *filtering* juga melakukan penghapusan *stopwords*. *Stopwords* adalah kata umum (*common words*) yang bisaanya muncul dalam jumlah besar dan dianggap tidak memiliki makna.

Pembobotan

Pembobotan TF-IDF merupakan pembobotan dengan penggabungan *term frequency* dan *inverse document frequency*, untuk menghasilkan bobot yang komposit untuk setiap *term* kata dalam setiap dokumen.

Klasifikasi

Dalam klasifikasi dapat menggunakan *Precision* dan *Recall*. Tujuannya untuk menentukan apakah data yang dipakai sudah memenuhi keinginan atau belum. *Precision* adalah tingkat ketepatan informasi yang dianggap relevan oleh pengguna dengan jawaban yang diberikan oleh sistem. *Recall* adalah tingkat keberhasilan system dalam menemukan kembali sebuah data yang dianggap relevan atau sesuai dengan pengguna.

K-fold Cross Validation

K-fold cross validation merupakan proses meminimalkan bias yang terkait dengan pengambilan sampel acak dari sampel data *training* dalam membandingkan akurasi prediksi dari dua atau lebih metode. Dalam *k-fold cross validation*, data awal dibagi secara acak menjadi k subset (*fold*). Pelatihan dan pengujian dilakukan sebanyak k kali. Pada iterasi i, data subset, digunakan sebagai dataset pengujian, dan partisi yang tersisa secara kolektif digunakan untuk melatih model agar mendapatkan suatu model klasifikasi yang nantinya digunakan dalam proses pengujian. Model klasifikasi dilatih dan diuji k kali. Setiap kali dilatih pada semua kecuali satu *fold* dan diuji pada *fold* tunggal yang tersisa.

Analisis Sentiment

Analisis sentiment merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam kalimat opini

Naïve Bayes Classifier

Naïve Bayes Classifier merupakan sebuah algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger, 2007). *Naïve Bayes classifier* mengasumsikan bahwa ada atau tidaknya suatu fitur tertentu pada sebuah kelas tidak akan mempengaruhi keberadaan fitur yang lain.

Persamaan dari teorema *bayes* adalah :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Dimana :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik
- $P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- $P(H)$: Probabilitas hipotesis H (prior probabilitas)
- $P(X|H)$: Probabilitas X berdasar kondisi pada hipotesis H.
- $P(X)$: Probabilitas X

Evaluasi

Evaluasi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performansi dari sistem yang telah dibuat. Metode pengukuran yang digunakan adalah confusion matrix (classification matrix or a contingency table).

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 1. Matriks *Contingency* Prediksi dan Aktual

Dimana :

- a TP (*True Positives*): kelas yang diprediksi positif dan diprediksi oleh sistem klasifikasi kelas positif.
- b TN (*True Negatives*): kelas yang diprediksi negatif dan diprediksi oleh sistem klasifikasi kelas negatif.
- c FP (*False Positives*): kelas negatif tetapi diprediksi oleh sistem klasifikasi kelas positif.
- d FN (*False Negatives*): kelas positif tetapi diprediksi oleh sistem klasifikasi kelas negatif.

Ada beberapa dalam formula perhitungan performa klasifikasi salah satunya yaitu nilai akurasi biasa ditampilkan dalam presentase. Akurasi adalah nilai ketepatan dimana pengguna memprediksi suatu kata sesuai dengan jawaban suatu sistem. Berikut perhitungan nilai akurasi :

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

Wordcloud

Wordcloud merupakan pengembangan dari situs jejaring sosial berbasis *web* atau gambaran visual dari tabulasi frekuensi kata-kata dalam setiap bahan tertulis yang dipilih, seperti catatan kuliah, teks dalam buku atau sebuah situs internet.

3. METODE PENELITIAN

Dalam penelitian ini data yang digunakan adalah data sekunder. Data diambil dari Twitter dengan kata kunci *New Normal*. Populasi pada penelitian ini adalah seluruh tweets oleh pengguna *Twitter* mengenai *New Normal* dalam Bahasa Indonesia. Sampel yang digunakan dalam penelitian ini adalah data *tweet* sebanyak 3000 *tweet*. Teknik pengumpulan data berasal dari *tweets* dengan cara melakukan scraping menggunakan API (*Application Programming Interface*).

Langkah-langkah analisis data yang digunakan dalam penelitian ini sebagai berikut :

1. Melakukan *scraping* data menggunakan *Twitter* API.
2. *Text Preprocessing*
 - a. *Case Folding*

Proses penyamaan *case* dalam sebuah dokumen yang dilakukan untuk mempermudah pencarian. Semua karakter huruf pada *tweet* diubah menjadi huruf kecil menggunakan fungsi *tolower*.
 - a. *Tokenizing*

Proses pembuatan tokenisasi dengan menggunakan fungsi *gsub* untuk menghilangkan *mention*, URL, tanda baca, dan angka pada *tweet*.
 - b. *Filtering*

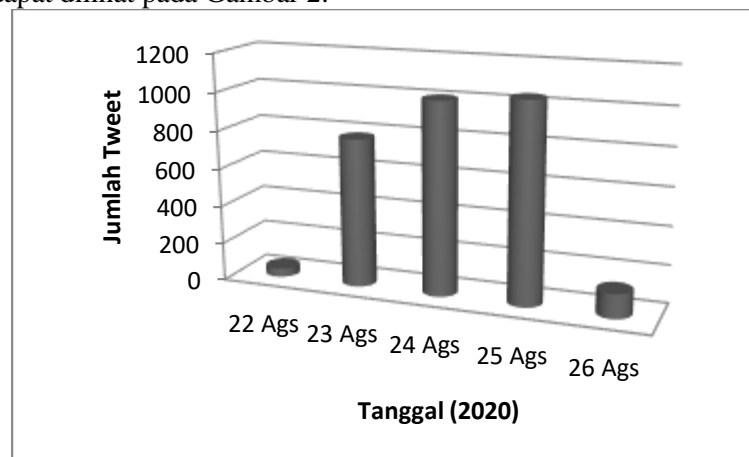
Pada tahap ini dilakukan pengambilan kata-kata penting dari hasil token menggunakan algoritma *stoplist* (membuang kata yang kurang penting).
3. Melakukan seleksi dengan TF-IDF yaitu mengubah teks menjadi vektor dan pembobotan kata.
4. Melakukan pembagian data menjadi data *training* dan data *testing* dengan *20-fold Cross Validation*.
5. Klasifikasi data menggunakan metode *Naïve Bayes Classifier* untuk menentukan hasil akurasi.

6. Visualisasi menggunakan *Wordcloud*, dalam tahap ini akan terbentuk kata-kata dari data *tweets* yang terdapat dalam *Twitter*.
7. Interpretasi dan mengambil kesimpulan.

4. HASIL PENELITIAN

Analisis Deskriptif

Analisis deskriptif dalam penelitian ini bertujuan untuk melihat gambaran secara umum tanggapan pengguna Twitter mengenai *New Normal* dengan melihat *tweet* yang ditulis yang sebelumnya diperoleh dengan teknik scraping. Setelah mendapatkan 3000 *tweet*, maka untuk melihat banyaknya orang yang membicarakan *New Normal* di *Twitter* per harinya dapat dilihat pada Gambar 2.



Gambar 2. Diagram Jumlah *Tweet*

Berdasarkan diagram di atas, 3000 tweet terbagi menjadi 5 hari. Jika dilihat dari diagram *tweet* selama 5 hari diperoleh hasil yang fluktuasi. Sebenarnya *tweet* pada 22 Agustus dan 26 Agustus lebih dari yang tertera pada diagram, namun karena pengambilan *tweet* dibatasi hanya untuk 3000 tweet terakhir sehingga hal tersebut dapat terjadi.

Text Preprocessing

a Case Folding

Proses ini dilakukan untuk penyeragaman bentuk huruf atau kata kedalam bentuk huruf kecil yang berguna agar huruf besar dan huruf kecil tidak terdeteksi memiliki perbedaan arti. Hasil dari proses *case folding* dapat dilihat pada Tabel 1.

Tabel 1. Hasil *Case Folding*

Sebelum	Sesudah
Eh tunggu dulu ada alasan ketiga, semenjak pandemi semuanya pada online tuh, sekolah online, kuliah online, kerja online https://t.co/Nl3AeExdA4	eh tunggu dulu ada alasan ketiga, semenjak pandemi semuanya pada online tuh, sekolah online, kuliah online, kerja online https://t.co/nl3aeexda4

b Tokenizing

Proses ini dilakukan untuk mendapatkan potongan kata atau token yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. Hasil dari proses *tokenizing* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Tokenizing*

Sebelum	Sesudah
eh tunggu dulu ada alasan ketiga, semenjak pandemi semuakan pada online tuh, sekolah online, kuliah online, kerja online https://t.co/nl3aeexda4	eh ; tunggu ; dulu ; ada ; alasan ; ketiga ; semenjak ; pandemi ; semuakan ; pada ; online ; tuh ; sekolah ; online ; kuliah ; online ; kerja ; online ; https://t.co/nl3aeexda4

c Filtering

Proses ini dilakukan untuk mengambil kata-kata penting dari hasil token dan pembersihan *tweet* dari spesial karakter, *URL link*, *username*, serta *emoticon* dan kata-kata lain. Hasil dari proses *filtering* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Filtering

Sebelum	Sesudah
eh ; tunggu ; dulu ; ada ; alasan ; ketiga ; semenjak ; pandemi ; semuakan ; pada ; online ; tuh ; sekolah ; online ; kuliah ; online ; kerja ; online ; https://t.co/nl3aeexda4	tunggu ; alasan ; ketiga ; semenjak ; pandemi ; online ; sekolah ; online ; kuliah ; online ; kerja ; online

Pelabelan Kelas Sentimen

Proses pelabelan pada data ini dilakukan secara otomatis dengan cara menghitung nilai pelabelan sentimen menggunakan kamus Lexicon dan manual. Pelabelan dibagi menjadi dua kelas sentimen, yaitu sentimen positif dan sentimen negatif dengan cara melakukan skoring. Penilaian dokumen ditentukan dengan memanfaatkan kumpulan kata dengan bahasa Indonesia yang terdiri dari kumpulan kata-kata positif dan kumpulan kata-kata negatif. Kemudian akan dihitung skor jumlah kata positif dikurangi dengan skor jumlah kata negatif dalam suatu kalimat ulasan. Jika suatu kalimat memiliki skor > 0 dan $= 0$ akan diklasifikasikan dalam kelas positif sedangkan jika kalimat memiliki skor < 0 diklasifikasikan dalam kelas negatif. Contoh hasil pelabelan data ulasan dapat dilihat pada Tabel 4.

Tabel 4. Contoh Hasil Pelabelan Data Ulasan

Kelas sentiment	Skor	Ulasan
Positif	2	melatih disiplin menerapkan peraturan hidup new normal hidup bersih
Negatif	-1	new normal mengabaikan protokol kesehatan pengelolanya gemes

Dalam penelitian ini kalimat ulasan yang masuk dalam kelas sentimen positif sebanyak 2409 ulasan sedangkan kalimat ulasan yang masuk dalam kelas negatif sebanyak 591 ulasan.

Data Training dan Testing

Data *training* digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*, semakin besar data *training* yang digunakan, maka akan semakin baik *machine* dalam memahami pola data. Data *testing* digunakan untuk mengukur sejauh mana

classifier berhasil melakukan klasifikasi dengan benar. Data yang digunakan untuk data *training* dan data *testing* adalah data yang telah memiliki label. Berdasarkan *Pareto Principle* rasio yang umum digunakan adalah 80:20 untuk data *sets training* dan *testing*. Perbandingan jumlah data latih dan data uji dapat dilihat pada Tabel 5:

Tabel 5. Perbandingan Data *Training* dan Data *Testing*

Klasifikasi	Jumlah	Data <i>Training</i> (80%)	Data <i>Testing</i> (20%)
Positif	2409	1.927,2 \approx 1927	481,8 \approx 482
Negatif	591	472,8 \approx 473	118,2 \approx 118
Total	3000	2400	600

Naive Bayes Classifier

Dalam klasifikasi menggunakan Naïve Bayes Classifier proses evaluasi untuk mengetahui hasil akurasi menggunakan *Confution Matrix* yang diperoleh hasil sebagai berikut :

Tabel 6. Confition Matrix data Tweet

	Aktual			Jumlah
	Class	Positif	Negatif	
Prediksi	Positif	446	34	480
	Negatif	42	78	120
Jumlah		488	112	

Hasil perhitungan dengan *Confition Matrix* didapatkan hasil prediksi dari 112 ulasan yang masuk dalam kelas negatif terdapat 78 ulasan yang telah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebanyak 34 ulasan yang masuk dalam ulasan positif. Sedangkan dari 488 ulasan yang masuk dalam kelas positif terdapat 446 ulasan terklasifikasi dengan benar dan terdapat kesalahan prediksi sebanyak 42 yang masuk dalam ulasan negatif. Dari hasil *confution matrix* tersebut diperoleh tingkat akurasi sebesar 87,33% yang artinya dari 600 data ulasan yang diuji terdapat 524 ulasan yang benar diklasifikasikan oleh *Naïve Bayes Classifier*.

Word Cloud

Pada tahap ini didapatkan hasil berupa infografis. Infografis yang berisikan kata-kata yang sering muncul pada data teks.

- Afshoh, F. 2017. "Analisa Sentimen Menggunakan Naïve Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter". Skripsi Sarjana Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta, Surakarta.
- Buntoro, G. A. 2017. "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter." *Integer Journal*, Vol 2, No 1 Hal: 32-41.
- Fahmi, A. (2020). *New Normal pada Media Sosial Twitter*. Retrieved from <https://iaingorontalo.ac.id/berita-821-new-normal-pada-media-sosial-twitter.html>
- Fari, K. (2020). *Twitter Sentiment Analysis during New Normal in Indonesia*. Retrieved from <https://medium.com/@muh.khoirulfajri.mkf/twitter-sentiment-analysis-during-new-normal-in-indonesia-ddab5896196f>
- Gumilang, Z.A.N. 2018. "Implementasi Naïve Bayes Classifier dan Asosiasi untuk Analisis Sentimen Data Ulasan Aplikasi E-Commerce Shopee pada Situs Google Play". Skripsi Sarjana Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia, Yogyakarta.
- Harijianto, S.D. 2019. "Analisis Sentimen pada Twitter Menggunakan Multinomial Naive Bayes". Skripsi Sarjana Fakultas Sains dan Teknologi Universitas Sanata Dharma, Yogyakarta.
- Kassambara, A. (2020). *Text mining and word cloud fundamentals in R : 5 simple steps you should know*. Retrieved from <http://www.sthda.com/english/wiki/text-mining-and-word-cloud-fundamentals-in-r-5-simple-steps-you-should-know>
- Pamungkas, D.S., Setiyanto N.A., Dolphina E. 2015. "Analisis Sentiment Pada Sosial Media Twitter Menggunakan Naive Bayes Classifier Terhadap Kata Kunci 'Kurikulum 2013'." *Techno. Com*, 14 (4), 299-314.
- Pratama, N.D., Sari, Y.A., Adikara P.P. 2018. "Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Chi Square Untuk Rekomendasi Lokasi Makanan Tradisional." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 9, hlm. 2982-2988.
- Putra, R.S. 2017. "Analisis Sentimen Twitter Dengan Klasifikasi Naïve Bayes Menggunakan Seleksi Fitur Mutual Information Dan Inverse Document Frequency". Skripsi Sarjana Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor, Bogor.
- Ratnawati, F. 2018. "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter." *Jurnal Inovtek Polbeng-Seri Informatika*, Vol.3, No.1.
- Rohit, K. (2016). *Naive Bayes Classification for Sentiment Analysis of Movie Reviews*. Retrieved from <https://rpubs.com/cen0te/naivebayes-sentimentpolarity>
- Syadid, F. 2019. "Analisis Sentimen Komentar Netizen terhadap Calon Presiden Indonesia 2019 dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (TF-IDF) dan Metode Multi Layer Perceptron (MLP) Neural Network". Skripsi Sarjana Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah, Jakarta.
- Syakuro, A. 2017. Analisis Sentimen Masyarakat terhadap E-Commerce pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) dengan Seleksi Fitur Information Gain (IG). Skripsi Sarjana Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim, Malang
- Syarifah, L. (2019). *Cara Cleaning Data Text Menggunakan Package tm di Rstudio*. Retrieved from <https://lailatussyarifah32.wordpress.com/2019/02/19/menguak-cara-explorasi-text-text-mining-dengan-package-tm-di-rstudi-statistics02/>
- We Are Social. (2020). *Data Internet di Indonesia dan Perilakunya Tahun 2020*. Retrieved from <https://teknoia.com/data-internet-di-indonesia-dan-perilakunya-880c7bc7cd19>

We Are Social. (2020). *Data Internet di Indonesia dan Perilakunya Tahun 2020*. Retrieved from <https://www.ayosemarang.com/read/2020/04/27/56031/saat-wabah-covid-19-social-media-marketing-jadi-pilihan-pemasaran-online>