

## **Analisis Sentimen Evaluasi Pembelajaran Tatap Muka 100 Persen pada Pengguna Twitter menggunakan Metode Logistic Regression**

**Saiful Anwar Assaidi<sup>1</sup>, Fatkhul Amin<sup>2</sup>**

<sup>1,2</sup> Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank

e-mail: saifulassa@gmail.com<sup>1</sup>, atkhulamin@edu.unisbank.ac.id<sup>2</sup>

### **Abstrak**

Perkembangan *Machine learning* semakin pesat seiring dengan bertambah banyaknya data yang ada di internet terutama dari media sosial seperti *facebook*, *twitter* dan *Instagram*. Diantara banyak topik yang dibahas di media sosial, topik tentang Evaluasi Pembelajaran Tatap Muka (PTM) seratus persen di Indonesia sempat hangat. Banyak masyarakat yang berpendapat mengenai evaluasi PTM terutama di media sosial twitter. banyaknya data yang tidak teratur dan berbeda pola serta makna sehingga sulit mengetahui pendapat yang disampaikan tersebut termasuk pendapat positif yang pro terhadap PTM atau sebaliknya, negatif yang kontra terhadap PTM. *Machine Learning* bisa mengolah data yang banyak dan tidak teratur untuk menghasilkan sebuah output dengan cepat. Tujuan penelitian ini agar menghasilkan sebuah *model* yang dapat membantu mengklasifikasi opini masyarakat terkait topik evaluasi Pembelajaran Tatap Muka (PTM) seratus persen kedalam sentiment negatif atau positif, biasa disebut dengan analisis sentimen. Klasifikasi termasuk dalam jenis *supervised learning* dalam *machine learning*. *Logistic Regression* merupakan metode *machine learning* yang dapat digunakan untuk membuat model Klasifikasi yang bisa digunakan untuk melakukan analisis sentimen. Hasil dari Analisis Sentimen menggunakan metode *Logistic Regression* pada penelitian ini menghasilkan nilai *accuracy* sebesar 78,57%, *precision* 76,92%, *recall* 83,3 % dan nilai *F1-Score* sebesar 80%.

**Kata kunci:** *Machine Learning*, Analisis Sentimen, *Twitter*, PTM, *Logistic Regression*

### **Abstract**

The development of machine learning is growing rapidly along with the increasing amount of data on the internet, especially from social media such as Facebook, Twitter and Instagram. Among the many topics discussed on social media, the topic of one hundred percent Face-to-Face Learning Evaluation (PTM) in Indonesia was hot. Many people think about the evaluation of PTM, especially on Twitter social media. the amount of data that is irregular and has different patterns and meanings so it is difficult to know the opinions expressed include positive opinions that are pro against PTM or vice versa, negative opinions against PTM. Machine Learning can process large and irregular data to produce an output quickly. The purpose of this research is to produce a model that can help classify public opinion related to the one hundred percent Face-to-Face Learning (PTM) evaluation topic into negative or positive sentiment, commonly referred to as sentiment analysis. Classification is included in the type of supervised learning in machine learning. Logistic Regression is a machine learning method that can be used to create a Classification model that can be used to perform sentiment analysis. The results of the Sentiment Analysis using the Logistic Regression method in this study resulted in an accuracy value of 78.57%, precision 76.92%, recall 83.3% and the F1-Score value of 80%.

**Keywords :** *Machine Learning*, Analisis Sentimen, *Twitter*, PTM, *Logistic Regression*

## PENDAHULUAN

*Machine Learning* telah berkembang pesat beberapa tahun terakhir. perkembangannya seiring dengan bertambah banyaknya data yang ada di internet terutama media sosial seperti *facebook*, *twitter* dan *Instagram* yang merupakan sumber data yang sangat besar. Sulit untuk mengubah data *input* yang begitu banyak menjadi suatu *output* dengan algoritma biasa, karena hubungan dan pola antara *input* dan *output* belum diketahui. Dengan *machine learning* data yang sangat besar dan tidak teratur dapat diproses dan dianalisis dengan waktu yang singkat. Misalnya dalam kasus klasifikasi data, *Machine learning* hanya perlu diberi contoh *input* dan *output* ke dalam algoritmanya sehingga bisa menghasilkan model yang dapat dipakai dalam klasifikasi.

Dua tahun sudah pandemi *COVID-19* melanda dunia tak terkecuali Indonesia. Namun kasus terinfeksi masih sering naik turun diberbagai negara termasuk indonesia. Berbagai kebijakan sudah dikeluarkan oleh Pemerintah Indonesia dan hasilnya sempat ada penurunan trend positif dan kasus aktif yang terjadi pada akhir tahun 2021. Hal ini tentu telah berdampak positif pada banyak bidang tak terkecuali bidang Pendidikan yang sudah beberapa bulan terakhir menerapkan PTM (Pembelajaran Tatap Muka) terbatas, setelah setahun lebih menerapkan PJJ (Pembelajaran Jarak Jauh) atau pembelajaran daring di berbagai daerah. Awal tahun 2022 Pemerintah memberlakukan PTM 100 Persen untuk wilayah Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) level 1-2 pada Januari 2022[1]. Namun masih banyak pihak yang menyayangkan dan kontra terhadap pemberlakuan PTM 100 persen tersebut karena vaksinasi untuk anak usia 6 - 11 tahun masih belum maksimal dan ditambah lagi dengan adanya kasus varian baru yang terdeteksi telah masuk ke Indonesia pada akhir tahun 2021 yaitu varian *omicron*. Setelah beberapa minggu diberlakukan PTM 100 persen di beberapa daerah, kasus penularan kian melonjak kembali pada akhir bulan januari karena adanya varian baru *omicron*. Menyusul dengan bermunculannya kasus baru dari kluster sekolah dan keluarga, semakin banyak pihak berharap penerapan PTM 100 persen agar segera dihentikan. Tapi banyak juga yang masih ingin melanjutkan PTM. Banyak respon di sosial media terutama berupa *tweet* di *twitter* yang diberikan terkait PTM yang belum diketahui merupakan golongan yang pro atau kontra terhadap kebijakan PTM karena tidak mengandung makna yang sama, sehingga dibutuhkan suatu model *machine learning* yang dapat menentukan pendapat tersebut masuk ke dalam klasifikasi pro atau kontra.

Analisis Sentiment merupakan cara mengumpulkan pendapat orang banyak dan menentukan sebuah dataset berbasis teks dari media sosial yang akan menghasilkan kelas positif dan negatif. Untuk melakukan analisis sentiment ada beberapa algoritma yang bisa digunakan. *Logistic Regression* merupakan salah satu algoritma untuk klasifikasi dalam machine learning dengan *output* berupa kelas positif dan negatif yang dapat digunakan untuk melakukan analisis sentiment. Penelitian ini bertujuan untuk mengetahui bagaimana sentimen masyarakat terhadap evaluasi pembelajaran tatap muka 100 persen selama akhir januari hingga awal february tahun 2022 seiring meningkatnya kasus terinfeksi virus Covid-19 terutama karena adanya varian baru yaitu *omicron* melalui *tweet* yang ada pada media sosial *twitter* menggunakan algoritma *Logistic Regression*.

Penelitian Antinasari, P., Perdana, R., & Fauzi, M, "Analisis Sentimen Tentang Opini Film pada Dokumen Twitter Berbahasa Indonesia Menggunakan *Naive Bayes* Dengan Perbaikan Kata Tidak Baku" (2017), Yang melakukan penelitian menggunakan kamus kata tidak baku dan normalisasi *Levenshtein Distance* untuk memperbaiki kata yang tidak baku menjadi kata baku dengan pengklasifikasian *Naive Bayes*. Berdasarkan hasil pengujian yang telah dilakukan didapatkan akurasi tertinggi dengan nilai *accuracy*, *precision*, *recall*, dan *f-measure* sebesar 98.33%, 96.77%, 100%, dan 98.36%[14].

Penelitian yang dilakukan Edward Darmaja, Viny Christanti Mawardi, Novario Jaya Perdana, "Review Sentimen Analisis Aplikasi Sosial Media di *Google Playstore* Menggunakan Metode *Logistic Regression*" (2021), Yang menggunakan TF-IDF dan *logistic regression* untuk membantu klasifikasi dari data *review* yang telah dikumpulkan. Dari hasil

penelitian tersebut menyatakan bahwa metode *TF-IDF* dan *logistic regression* dapat membantu dalam klasifikasi review kedalam sentiment positif maupun negative[10].

Penelitian Andry Novantika, Sugiman, "Analisis Sentimen Ulasan Pengguna Aplikasi Video Conference Google Meet menggunakan Metode SVM dan Logistic Regression" (2022) , yang menggunakan data dari hasil scrapping sebanyak 5160 data dari ulasan Google Play Store terhadap aplikasi Zoom Meeting yang terdiri dari 2786 ulasan positif dan 2374 ulasan negatif. Didapatkan hasil nilai akurasi sebesar 85,17% untuk metode logistic regression dan 87,02% untuk metode SVM[18].

### **Pembelajaran Tatap Muka (PTM)**

Pembelajaran Tatap Muka merupakan proses interaksi antara guru dengan siswa yang terjadi secara langsung dan konvensional tanpa menggunakan media *online*. PTM akan memberikan pengalaman belajar langsung kepada siswa untuk meningkatkan keberhasilan belajar, melalui interaksi yang tercipta ketika proses pembelajaran berlangsung antara guru dengan siswa.[2]

### **Twitter**

*Twitter* merupakan jejaring sosial dan *microblog* yang memungkinkan penggunaanya untuk bertukar pesan berbasis teks dengan batasan hingga 140 karakter, pesan tersebut yang dinamakan kicauan (*tweet*). *Twitter* sendiri didirikan oleh Jack Dorsey pada bulan Maret 2006 sementara pada bulan Juli 2006 jejaring sosialnya baru diluncurkan[3]. Menurut data *Statista*, ada 18,45 juta pengguna twitter di indonesia per Januari 2022 lalu, Sehingga menempatkan Indonesia di posisi ke 5 sebagai negara dengan pengguna twitter terbanyak di dunia[4].

### **Machine Learning**

Teknik untuk melakukan inferensi terhadap data dengan pendekatan matematis. Inti *Machine Learning* adalah untuk membuat model (matematis) yang merefleksikan pola-pola data. Inferensi yang dimaksud lebih menitikberatkan ranah hubungan variabel. Misalnya, apakah penjualan meningkat apabila kita meningkatkan biaya marketing. Tujuan dari *machine learning* minimal ada dua, memprediksi masa depan dan/atau memperoleh ilmu pengetahuan. [6]

*Machine Learning* terdiri dari dua kategori besar yaitu *supervised machine learning* dan *unsupervised machine learning*. *Supervised machine learning* adalah teknik yang digunakan untuk mengetahui hubungan antara atribut input dengan atribut target. Klasifikasi merupakan salah satu kategori dari *supervised machine learning*. Klasifikasi adalah pendekatan data mining yang digunakan untuk memprediksi kategori grup untuk contoh data.[9][13]

### **Analisis Sentimen (Sentiment Analysis)**

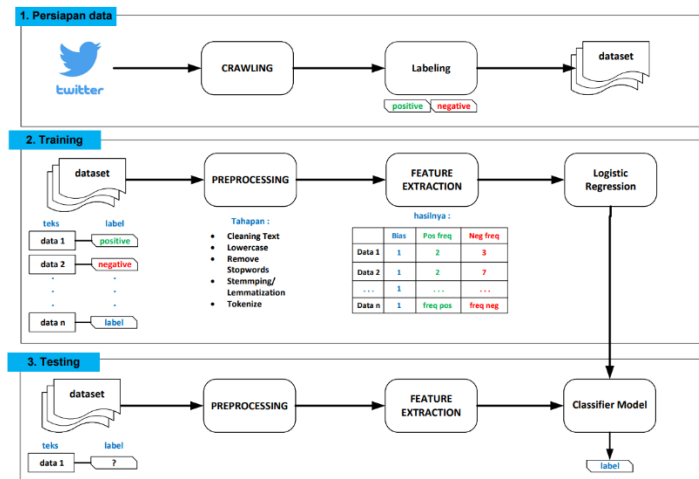
Analisis Sentiment adalah cara mengumpulkan pendapat orang banyak mengenai sebuah isu terkini menggunakan media sosial, yang dapat digunakan untuk mengevaluasi pelayanan, suatu kinerja, dan sebagainya.[11]

Menurut penelitian Bintang, Yuyun & Rini (2021) Menjelaskan bahwa Analisis Sentimen merupakan suatu proses yang menentukan sebuah dataset berbasis teks yang akan menghasilkan kelas positif atau negatif.[12]

### **Logistic Regression**

Merupakan metode *Supervised machine learning* yang dapat digunakan untuk menganalisis data serta mendeskripsikan antara satu atau lebih variabel prediksi dengan satu variabel respon. Variabel respon dari *Logistic Regression* hanya bernilai antara 0 dan 1 sehingga akan menghasilkan *class* sentiment positif dan negatif[19] dengan batasan diantara keduanya adalah nilai 0,5[17]

## METODE PENELITIAN



Gambar 1. Alur Penelitian.

### Persiapan Data

Persiapan data terdiri dari 2 proses yaitu *Crawling* dan *Labeling* yang menghasilkan *dataset*, dapat dilihat pada Gambar 1.

#### 1. *Crawling*

Merupakan salah satu tahap dalam penelitian dengan tujuan untuk mengunduh atau mengumpulkan data dari suatu database [8]. data yang diunduh adalah data dari *server twitter* berupa *user* dan *tweet* dengan menggunakan *library tweepy* dari bahasa pemrograman *python*. Contoh data hasil *crawling* ditunjukkan pada Tabel 1 berikut.

Table 1. Sampel hasil *Crawling*

Id	Text tweet	User
1	Bener enakan ptm Sebenarnya mah klo gada kasus meninggal yaa terobos aj sih PTM, org klo di PJJ juga pada keluyuran main	jnghwnuna_
2	sana sini	jnghwnuna_
3	Lebih males ptm noted! rada takut juga sih kalo mau sekolah pas ptm karena rame banget dan udah ngerasa bakal ada yang kena	kakuchodet
4	:( enakan ptm tau bisa ketemu temen +	callamily
5	DAPET UANG JAJAN	nzzjm0

#### 2. *Labelling*

Proses pemberian tag atau label sentiment berupa label "*Negative*" untuk tweet yang kurang setuju dengan PTM, dan label "*Positive*" untuk tweet yang setuju dengan adanya PTM. Pelabelan dilakukan secara manual menggunakan teks editor. Contoh data setelah terlabeli dapat dilihat pada tabel X berikut.

**Table 2. Sampel hasil Labelling**

Id	Sentiment	Text tweet	User
1	positive	Bener enakan ptm	jnghwnuna_
2	positive	Sebenarnya mah klo gada kasus meninggal yaa terobos aj sih PTM, org klo di PJJ juga pada keluyuran main sana sini	jnghwnuna_
3	negative	Lebih males ptm	kakuchodet
4	negative	noted! rada takut juga sih kalo mau sekolah pas ptm karena rame banget dan udah ngerasa bakal ada yang kena :(	callamily
5	positive	enakan ptm tau bisa ketemu temen + DAPET UANG JAJAN	nzzjm0

## Training

### 1. Preprocessing

Tahap pertama dalam training adalah *preprocessing*. Bertujuan untuk menyaring kumpulan data (*dataset*) agar berisi informasi yang penting saja untuk proses berikutnya. Hasilnya akan mengurangi ukuran kosa kata sehingga dapat menghemat waktu saat proses *training* dan *testing*[5]. Hasil dari proses preprocessing dapat dilihat pada Gambar 2.

#### a. Cleaning Text

Membersihkan teks dari tanda baca, tanda # (tagar) tanpa menghilangkan kata setelah # dan tanda @(mention).

#### b. Lowercase

Merubah semua huruf yang ada pada dataset menjadi huruf kecil semua.

#### c. Remove Stopword

Menghapus atau menghilangkan kata-kata yang tidak memiliki makna jika berdiri sendiri.

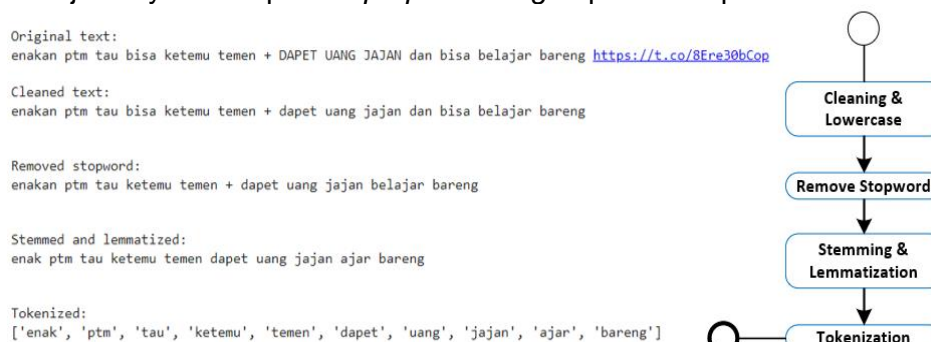
#### d. Stemming/Lemmatization

Menghapus imbuhan yang dimiliki oleh masing-masing kata dan merubahnya menjadi kata dasar.

#### e. Tokenize

Memisahkan masing-masing kata dalam semua teks menjadi token token kata. Masing-masing token kata tersebut hanya terdiri dari satu kata saja[15].

Lebih jelasnya untuk proses *preprocessing* dapat dilihat pada Gambar 2 berikut.



**Gambar 2. Contoh proses preprocessing**

## 2. Feature Extraction

Bertujuan untuk merepresentasikan tesk menjadi vektor 3 dimens berdasarkan seberapa sering sebuah kata muncul dalam kategori sentimen tertentu (positif atau negative) berdasarkan *Vocabulary* yang sudah dibuat oleh sistem [5]. Rumus Feature extraction dapat dilihat pada persamaan 1 berikut.

$$X_n = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)] \quad (1)$$

## 3. Logistic Regression

Persamaan awal dari *logistic regression* adalah model linier untuk *binary classification* yang dapat dilihat pada persamaan 2 berikut.

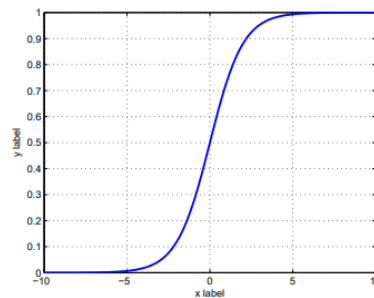
$$f(x) = x \cdot w + b \quad (2)$$

Persamaan 2 merepresentasikan suatu data sebagai *vector*  $x$  dan ada *bias*  $b$ . fungsi akan menghitung bobot setiap fitur pada *vector*  $x$  dengan mengalikannya (*dot product*) dengan parameter. Klasifikasi akan dilakukan dengan memasukan data pada fungsi yang mempunyai parameter. Persamaan 2 dapat ditulis juga seperti persamaan 3 berikut, dimana  $x_i$  merupakan elemen ke- $i$  dari *vector*  $x$ [6].

$$f(x) = b + w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n \quad (3)$$

Untuk membuat hasilnya berupa angka probabilitas yang bernilai antara 0 dan 1 diperlukan *sigmoid function*. Dari persamaan (2) nilai  $z$  akan dimasukan kedalam *sigmoid function*, ( $W$ ). Bernama *Sigmoid function* (karena hasilnya akan membentuk diagram seperti huruh s) lihat Gambar 3 atau bisa disebut juga *logistic function*. Maka persamaannya akan seperti berikut.

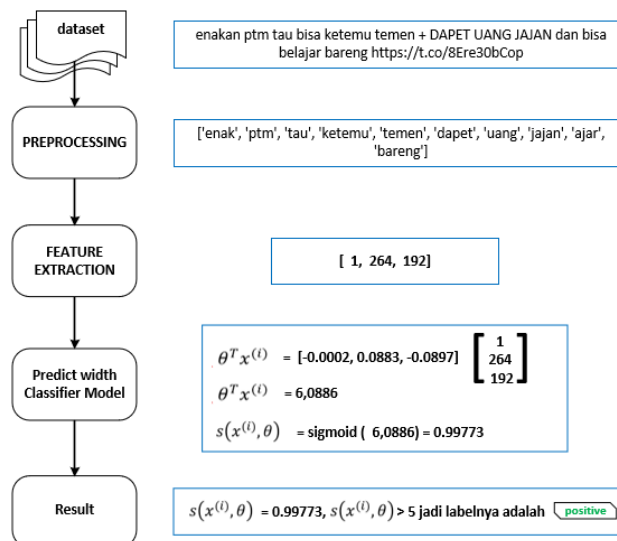
$$y = \sigma(f(x)) = \frac{1}{1 + e^{-f(x)}} \quad (4)$$



Gambar 3. Diagram *sigmoid function*[7].

## Testing

Proses testing dimulai dengan memasukan data berupa teks *tweet* yang masih utuh dengan tanda baca dan atribut lainnya. Kemudian akan dilakukan *preprocessing* dan *feature extraction* oleh model yang sudah dibuat. Hasil dari *feature extraction* akan dijadikan bahan untuk melakukan prediksi oleh *classifier model* dengan mengalikan hasil *feature extraction* tadi dengan *weight theta* yang diperoleh dari proses training. Lebih jelasnya proses testing dapat dilihat pada Gambar 4.



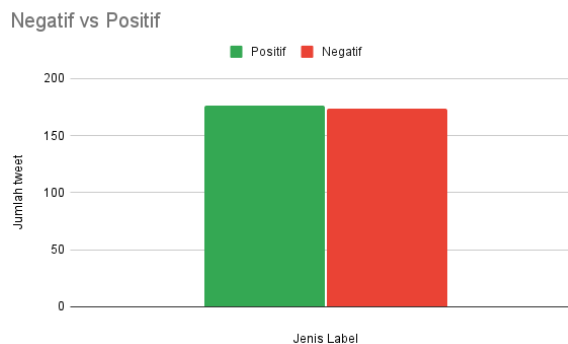
**Gambar 4. Ilustrasi proses testing.**

Hasil dari Gambar 4 menunjukkan bahwa hasilnya berupa nilai 0,99773 atau nilainya lebih dari 5, maka label untuk data tersebut akan dimasukkan kedalam kategori positif. Karena jika nilai > 0,5 maka akan dimasukkan ke dalam kategori positif, sementara jika hasil nilainya < 0,5 akan dimasukkan ke dalam kategori negative.

## HASIL DAN PEMBAHASAN

### Hasil Penelitian

Penelitian ini menggunakan dataset dari *tweet* terkait topik evaluasi PTM 100% dengan dataset berjumlah 350 tweet yang terdiri dari 176 tweet positif dan 174 tweet negatif dapat dilihat perbandingan pada Gambar 5.



**Gambar 5. Perbandingan jumlah tweet positif dan negatif.**

Sampel 5 data dari dataset ditunjukkan pada Tabel 3 berikut .

**Table 3. Sampel data dari dataset**

id	Sentiment	Text Tweet
1	negative	ptm 50% gaenak Ya Allah :(
2	positive	enakan ptm Ya Allah ketemu ginian lagi, ya kalo sekolahnya pjj ya udah pjj, kalo sekolahnya ptm yaudah ptm dah. Gw sih tim ptm \n\nKatanya kita harus hidup berdampingan dengan virus, mau ga mau hidup
3	positive	atau ngga ya tetep ae hidup berjalan ngalir'
4	positive	enak msi ptm
5	positive	semangat ptm nya nia

Wordcloud dari sentiment positif dan negatif dapat dilihat pada gambar 5 berikut.



**Gambar 6. Wordcloud sentimen positif dan negatif**

Dataset yang sudah disiapkan tadi akan diproses dalam tahap *preprocessing*. Hasil dari tahapan tersebut dapat dilihat pada sampel 5 data yang sudah diproses pada tabel 4 berikut.

**Table 4. Sampel hasil *preprocessing*.**

id	Sentiment	Text Tweet
1	<i>negative</i>	['ptm', '50', 'gaenak', 'allah']
2	<i>positive</i>	['enak', 'ptm']
3	<i>positive</i>	['allah', 'ketemu', 'ginian', 'sekolah', 'pjj', 'pjj', 'sekolah', 'ptm', 'yaudah', 'ptm', 'gw', 'tim', 'ptm', 'n', 'nkatanya', 'damping', 'virus', 'ga', 'ngga', 'tete p', 'ae', 'ngalir']
4	<i>positive</i>	['enak', 'msi', 'ptm']
5	<i>positive</i>	['semangat', 'ptm', 'nia']

Dataset yang sudah melalui *preprocessing* akan dibagi menjadi data *training* dan data *testing dengan* perbandingan 80 : 20. Berikut ditunjukkan sampel 5 data yang dijadikan sebagai data *testing* pada gambar 7.

145	[enak, ptm, maju, ngerjain, nanya, bagus, dape...
157	[ptm, fuck, daring]
89	[anak, ajar, mau, ptm, ulang, mau, online]
287	[lbh, suka, ptm, jujur, rumah, kaya, ga, kerja...
78	[plis, sekolah, ptm, gajadi]

**Gambar 7. Sampel data *testing*.**

Data testing akan diprediksi oleh *Classifier Model* untuk memprediksi sentiment dengan *output* berupa nilai diantara 0 sampai dengan 1. Lebih jelasnya hasil dari prediksi dengan *Classifier Model* dapat dilihat pada Gambar 8 berikut.

[[ 1. 191. 160.]]	['enak', 'ptm', 'maju', 'ngerjain', 'nanya', 'bagus', 'dapet', 'nilai', 'tambah'] -> 0.961114
[[ 1. 164. 156.]]	['ptm', 'fuck', 'daring'] -> 0.705734
[[ 1. 181. 185.]]	['anak', 'ajar', 'mau', 'ptm', 'ulang', 'mau', 'online'] -> 0.418352
[[ 1. 367. 351.]]	['lbh', 'suka', 'ptm', 'jujur', 'rumah', 'kaya', 'ga', 'kerja', 'kamar', 'mulu', 'bosen', 'pgn', 'ptm', 'lagiiiiii'] -> 0.853917
[[ 1. 183. 196.]]	['plis', 'sekolah', 'ptm', 'gajadi'] -> 0.224927
[[ 1. 227. 194.]]	['huhu', 'online', 'tuh', 'kdg', 'ga', 'anak', 'moga', 'cepat', 'ptm'] -> 0.968450

**Gambar 8. Sampel hasil prediksi.**

### Hasil Matriks Evaluasi

*Confusion Matrix* merupakan tabel matriks merepresentasikan performa *classifier model* pada jumlah hasil prediksi label dataset yang dijadikan sebagai data testing terhadap jumlah nilai label sebenarnya yang sudah diketahui pada dataset. Prediksi kelas positif yang benar dinamakan *True Positive* (TP), sementara yang salah dinamakan *False Positive*.



Untuk kelas negatif yang diprediksi benar diberi nama True Negative (TN), untuk yang salah dinamakan *False Negative* (FN). Hasil dari Confusion Matrix dapat dilihat pada Tabel 5 berikut.

Tabel 5. Hasil *Confusion Matric*

Label Sebenarnya	Hasil Prediksi Label	
	<i>Negative</i>	<i>Positive</i>
<i>Negative</i>	25 (TN)	9 (FN)
<i>Positive</i>	6 (FP)	30 TP)

1. Akurasi (*Accuracy*)

Merupakan tingkat kesamaan nilai hasil prediksi dengan nilai sebenarnya[19], atau tingkat keberhasilan model dalam melakukan prediksi dengan tepat. Nilai Akurasi dapat dihitung dengan persamaan X dan Y berikut.

$$Accuracy = \frac{\text{jumlah prediksi yang benar}}{\text{jumlah data test}} \quad (5)$$

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

2. *Precision*

Merupakan tingkat ketepatan prediksi model terhadap sentiment positif [21]. Dapat dihitung dengan persamaan X berikut

$$precision = \frac{TP}{TP+FP} \quad (7)$$

3. *Recall*

Nilai tingkat kesuksesan dalam mengenali kembali suatu kelas yang harus dikenali[16]. rumus untuk menghitung nilai recall ada pada persamaan X berikut.

$$recall = \frac{TP}{TP+FN} \quad (8)$$

4. *F1-Score*

Merupakan gabungan nilai dari *precision* dan *recall* atau nilainya bisa mewakili keseluruhan kinerja suatu model[20] . F1-Score dapat dihitung dengan persamaan berikut.

$$F1 = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \quad (9)$$

Rangkuman untuk semua perhitungan yang telah dilakukan dapat dilihat pada tabel 6 berikut.

Tabel 6. Hasil perhitungan evaluasi

Jenis Perhitungan	Accuracy	Precision	Recall	F1-Score
Hasil	78,57%	76,92%	83,3%	80%

**SIMPULAN**

1. Metode *Logistic Regression* dapat digunakan untuk membuat *classifier model* yang dapat membantu klasifikasi sentiment terhadap topik evaluasi Pembelajaran Tatap Muka (PTM) pada media sosial *twitter*.
2. Output yang dihasilkan dari *classifier model* dengan menggunakan metode *Logistic Regression* berupa nilai antara 0 sampai dengan 1 dengan Batasan diantara keduanya

berupa nilai 0,5, dimana jika nilai yang dihasilkan  $> 0,5$ , maka akan dimasukkan kedalam label positif (1), jika sebaliknya maka akan masuk kedalam label negatif (0).

3. Hasil evaluasi model yang didapat dari 349 *tweet* berupa 177 *tweet* positif dan 172 *tweet* negatif dengan pembagian antara data training dibandingkan data testing sebanyak 80:20 dihasilkan nilai *accuracy* sebesar 78,57%, *precision* 76,92%, *recall* 83,3 % dan nilai *F1-Score* sebesar 80%.

## DAFTAR PUSTAKA

- Zintan Prihatini (2022). Pembelajaran Tatap Muka 100 Perser Mulai Digelar, Ini Saran Epidemiolog, Sains, <https://www.kompas.com/sains/read/2022/01/05/080200223/pembelajaran-tatap-muka-100-persen-mulai-digelar-ini-saran-epidemiolog?page=all>, diakses tanggal 3 februari
- Kembang, L. G. (2020). *Perbandingan model pembelajaran tatap muka dengan model pembelajaran daring ditinjau dari hasil belajar mata pelajaran SKI (studi pada siswa kelas VIII) MTs Darul Ishlah Ireng Lauk Tahun Pelajaran 2019/2020* (Doctoral dissertation, Uin Mataram).
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng-Seri Informatika*, 3(1), 50-59.
- Annur, Cindy Mutia (2022). Pengguna Twitter Indonesia Masuk Daftar Terbanyak di Dunia, Urutan Berapa?, <https://databoks.katadata.co.id/datapublish/2022/03/23/pengguna-twitter-indonesia-masuk-daftar-terbanyak-di-dunia-urutan-berapa>, diakses tanggal 11 April 2022
- Chadha, Aman (2020). Sentiment Analysis using Logistic Regression, *Distilled Notes for the Natural Language Processing Specialization on Coursera (offered by deeplearning.ai)*, <https://aman.ai/coursera-nlp/logistic-regression/#feature-extraction>, diakses tanggal 1 Juni 2022
- Putra, Jan Wira Gotama. "Pengenal konsep pembelajaran mesin dan deep learning." Tokyo. Jepang (2019).
- Jin, Q., Zhang, Y., Cai, W., & Zhang, Y. (2020). A new similarity computing model of collaborative filtering. *IEEE Access*, 8, 17594-17604.
- Sembodo, J. E., Setiawan, E. B., & Baizal, Z. A. (2016, October). Data Crawling Otomatis pada Twitter. In *Indonesian Symposium on Computing (Indo-SC)* (pp. 11-16).
- Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied computing and informatics*, 15(1), 27-33.
- Darmaja, E., Mawardi, V. C., & Perdana, N. J. (2021). REVIEW SENTIMEN ANALISIS APLIKASI SOSIAL MEDIA DI GOOGLE PLAYSTORE MENGGUNAKAN METODE LOGISTIC REGRESSION. *PROSIDING SERINA*, 1(1), 513-520.
- Syarifuddin, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn. *Inti Nusa Mandiri*, 15(1), 23-28.
- Amalia, B. S., Umaidah, Y., & Mayasari, R. (2021). Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor. *SITEKIN: Jurnal Sains, Teknologi dan Industri*, 19(1), 28-34.
- Savitri, N. L. P. C., Rahman, R. A., Venyutzky, R., & Rakhmawati, N. A. (2021). Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter

- Menggunakan Supervised Machine Learning. *Jurnal Teknik Informatika dan Sistem Informasi*, 7(1).
- Antinasari, P., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1733–1741
- Habibi, R., Setyohadi, D. B., & Wati, E. (2016). Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation. *Jurnal Informatika*, 12(1), 103–109
- Rahutomo, F., Saputra, P. Y., & Fidyawan, M. A. (2018). Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. *Jurnal Informatika Polinema*, 4(2), 93. <https://doi.org/10.33795/jip.v4i2.152>
- Wan, X. (2019, May). The influence of polynomial order in logistic regression on decision boundary. In *IOP Conference Series: Earth and Environmental Science* (Vol. 267, No. 4, p. 042077). IOP Publishing.
- Novantika, A., & Sugiman, S. (2022, February). Analisis Sentimen Ulasan Pengguna Aplikasi Video Conference Google Meet menggunakan Metode SVM dan Logistic Regression. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 5, pp. 808-813).
- Bimantara, A., & Dina, T. A. (2019, May). Klasifikasi Web Berbahaya Menggunakan Metode Logistic Regression. In *Annual Research Seminar (ARS)* (Vol. 4, No. 1, pp. 173-177).