

Terbit online pada laman: <http://jurnal.utu.ac.id/JTI>

## Jurnal Teknologi Informasi

| ISSN (Print): xxx-xxx | ISSN (Online): xxx-xxx |



# Analisis Sentimen Cyberbullying Pada Komentar Instagram Dengan Metode Klasifikasi Naïve Bayes

Saifullah<sup>1</sup>, Marhamah<sup>2</sup>, Sofia<sup>3</sup>, Zainul Mutaqin<sup>4</sup>,

Universitas Teuku Umar

Jl. Alue Peunyareng, Ujong Tanoh Darat, Kec. Meureubo, Kabupaten Aceh Barat, Aceh 23681

Email : <sup>1</sup>saifullah.tif20@gmail.com, <sup>2</sup>marhamah1122003@gmail.com, <sup>3</sup>safia042002@gmail.com, <sup>4</sup>zainmttqin@gmail.com

## INFORMASI ARTIKEL

Sejarah Artikel:

Diterima:

Revisi:

Diterbitkan:

Kata Kunci:

Analisis Sentimen

Cyberbullying

Text Classification

Metode Naïve Bayes

## ABSTRAK

Instagram digunakan oleh pengguna untuk berbagi momen lewat foto, video. Dan juga Instagram memiliki fitur untuk mengomentari foto dan video pengguna orang lain. Fitur tersebut memiliki sisi positif dan sisi negative, sisi positif nya pengguna bisa merespon postingan orang lain dan sisi negative nya pengguna bisa melakukan tindak yang buruk seperti mengancam, merendahkan ataupun menjelek-jelekan penggun Instagram. Didalam dunia social media tindakan itu disebut dengan *Cyberbullying*. *Cyberbullying* adalah intimidasi melalui media elektronik. Penelitian ini bertujuan untuk membuat model mesin pembelajaran dengan menggunakan metode klasifikasi *naïve bayes* dalam kasus analisis sentimen pada komentar Instagram tersebut, guna untuk menyaring dan mengetahui komentar-komentar Instagram yang bersifat positif atau negatif. Dari hasil analisis sentiment yang didapatkan itu akan berguna untuk dilakukannya tindakan preventif baik untuk korban atau pelaku *cyberbullying*. Hasil dari metode klasifikasi *naïve bayes* memberikan nilai akurasi 91%, persisi 97%, recall 86%, dan F-1 Score 91%. Dengan menggunakan splitting atau pembagian data 80 20, 80% data training dan 20% data testing. Dan hasil dari performa masing-masing sentiment atau kategori memiliki nilai yang berbeda yaitu, sentimen negative memiliki nilai persisi 97%, recall 86%, dan F-1 Score 92%. Sedangkan sentiment positive memiliki nilai persisi 85%, recall 97%, dan F-1 Score 91%. Artinya dataset komentar Instagram yang digunakan pada penelitian ini memiliki nilai sentiment negative yang tinggi.

Copyright © 2022 Jurnal Teknologi Informasi UTU  
All rights reserved

## 1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi berkembang pesat, hal tersebut di dukung oleh peralatan *device* dan internet yang mudah di gunakan atau mudah di akses bagi masyarakat. Internet menjadi salah satu poin utama untuk menggunakan teknologi informasi dan komunikasi secara digital.

Berdasarkan survei dari (APJI) Asosiasi Penyelenggara Jasa Internet di Indonesia pada tahun 2022-2023, pengguna internet di Indonesia mencapai 215,63 juta. Jumlah tersebut meningkat 2,67% dibandingkan pada periode sebelumnya yang sebanyak 210,03 juta pengguna. Jumlah pengguna internet tersebut setara dengan 78,19% dari total populasi Indonesia sebanyak 275,77 juta. Para remaja dan dewasa mengakses internet untuk memperoleh informasi, hiburan dan juga berkomunikasi dengan orang-orang yang tidak mereka kenal ataupun menambah teman dengan menggunakan media social [1].

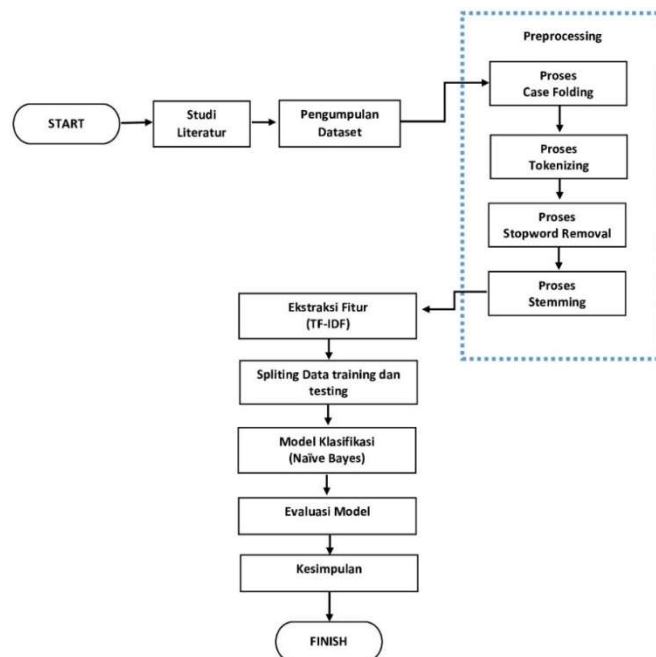
Instagram menjadi salah satu media social yang sangat populer saat ini. Pada survey yang dilakukan oleh Napoleon cat, ada 109,33 juta pengguna Instagram di Indonesia hingga april 2023. Jumlah tersebut meningkat 3,45% dibandingkan pada bulan sebelumnya yang sebesar 105,68 juta pengguna [2]. Instagram digunakan oleh pengguna untuk berbagi momen lewat foto, video. Dan juga Instagram memiliki fitur untuk mengomentari foto dan video pengguna orang lain. Fitur tersebut memiliki sisi positif dan sisi negative, sisi positif nya pengguna bisa merespon postingan orang lain dan sisi negative nya pengguna bisa melakukan tindak yang buruk seperti mengancam, merendahkan ataupun menjelek-jelekan penggun Instagram. Didalam dunia social media tindakan itu disebut dengan *Cyberbullying*.

*Cyberbullying* adalah intimidasi melalui media elektronik. Hal ini dapat terwujud dalam bentuk memposting rumor, ancaman, komentar seksual, informasi pribadi korban, atau komentar yang merendahkan. Orang-orang yang menggunakan social media akan rentan mendapatkan *cyberbullying* [3]. Permasalahan *Cyberbullying* pada kolom komentar Instagram menjadi hal yang penting untuk dikaji sebagai pemrosesan teks.

Penelitian ini bertujuan untuk membuat model mesin pembelajaran dengan menggunakan metode klasifikasi *naïve bayes* dalam kasus analisis sentimen pada komentar Instagram tersebut, guna untuk menyaring dan mengetahui komentar-komentar Instagram yang bersifat positif atau negatif. Dari hasil analisis sentiment yang didapatkan itu akan berguna untuk dilakukannya tindakan preventif baik untuk korban atau pelaku *cyberbullying*

## 2. Metodologi Penelitian

Adapun terdapat beberapa proses penelitian yang digambarkan melalui diagram alur penelitian berikut:



Gambar 1. Diagram alur penelitian

## 2.1. Studi Literatur

Studi literatur akan mengumpulkan data-data dari penelitian sebelumnya, salah satu penelitian sebelumnya yang menjadi referensi pada penelitian ini adalah penelitian yang telah dilakukan oleh Wanda Athira Luqyana, tentang Analisis Sentimen Cyberbullying pada komentar Instagram dengan metode klasifikasi support vector machine hasil penelitian tersebut menghasilkan nilai akurasi tertinggi sebesar 90% pada komposisi data latih 50% dan komposisi data uji 50% [4]. Pada penelitian ini akan membandingkan nilai akurasi dari penelitian sebelumnya dengan menggunakan metode yang berbeda yaitu metode klasifikasi Naïve Bayes.

## 2.2. Pengumpulan Dataset

Ruang lingkup pengumpulan dataset pada kasus penelitian ini adalah komentar pengguna Instagram. Dan dataset yang digunakan pada penelitian ini menggunakan dataset yang telah di *scarping* (mengumpulkan) dan dilabeling oleh penelitian sebelumnya [4]. Dataset tersebut memiliki total 400 data, yang dimana memiliki dua label yaitu positif dan negatif. Masing masing label tersebut dibagi menjadi 200 data.

	Id	Sentiment	Instagram Comment Text
0	1	negative	<USERNAME> TOLOL!! Gak ada hubungan nya kegug...
1	2	negative	Geblek lo tata...cowo bgt dibela2in balikan.....
2	3	negative	Kmrn termewek2 skr lengket lg duhhh kok labil ...
3	4	negative	Intinya kalau kesel dengan ATT nya, gausah ke ...
4	5	negative	hadewwww perempuan itu lg!!!!sakit jiwa,knp ha...

Gambar 2. Contoh 5 data teratas yang ada di dataset.

## 2.3. Preprocessing Data

*Preprocessing* data adalah tahapan yang wajib dilakukan setelah melakukan scraping data. *Preprocessing* bertujuan untuk membersihkan data text yang kotor, contoh dari data text yang kotor yaitu seperti text ini "<USERNAME> TOLOL!! Gak ada hubungan nya". Text tersebut harus dibersihkan atau dirapikan. Agar dataset tersebut dapat digunakan semaksimal mungkin untuk meningkatkan nilai akurasi model yang akan dibuat. Semakin bagus tahap *preprocessing* data maka akan meningkatkan kualitas hasil dari pembelajaran model yang akan dibuat [5]. Proses *preprocessing* yang digunakan pada penelitian ini adalah sebagai berikut :

1. Proses *Case folding* : proses *case folding* bertujuan untuk membuat huruf menjadi kecil semua atau lowercase agar semua huruf tersebut menjadi bentuk standar atau menjadi sama semua. Contoh huruf: UANG menjadi uang.

	Id	Sentiment	Comment
0	1	negative	<username> tolol gak ada hubungan nya kegugura...
1	2	negative	geblek lo tatacowo bgt dibela2in balikanhadeww...
2	3	negative	kmrn termewek2 skr lengket lg duhhh kok labil ...
3	4	negative	intinya kalau kesel dengan att nya gausah ke a...
4	5	negative	hadewwww perempuan itu lgsakit jiwaknp harus d...
...	...	...	...
95	96	negative	yg nama ny plagiat ya plagiat walaupun laku tp...
96	97	negative	dasar plagiat semuanya ga bisa kreatifmemaluka...
97	98	negative	dan dengan begonya dia bilang apanya yang sama...
98	99	negative	itu pacarnya angel pietrs kali dan cowok itu ...
99	100	negative	japok gateeel emanguda jelas dia ada main sama...

Gambar 3. Hasil dataset yang telah di *Case folding*.

2. Proses *Tokenizing* : proses *tokenizing* bertujuan untuk membuat setiap kata menjadi berurutan dari 1 kalimat dan memisahkan setiap kata-kata tersebut dengan tanda koma di dalam satu kalimat. Contoh kalimat nya : “aku pergi ke pasar membeli buah” menjadi “[aku, pergi, ke, pasar, membeli, buah]”



	Id	Sentiment	Comment
0	1	negative	[<username>, tolol, gak, ada, hubungan, nya, k...
1	2	negative	[geblek, lo, tatacowo, bgt, dibela2in, balikan...
2	3	negative	[kmrn, termewek2, skr, lengket, lg, duhhh, kok...
3	4	negative	[intinya, kalau, kesel, dengan, att, nya, gaus...
4	5	negative	[hadewwww, permipuan, itu, lgsakit, jiwaknp, h...
5	6	negative	[pantesan, di, tinggalin, laki, ya, lakinya, j...
6	7	negative	[kebiasaan, balajaer, nyampah, d, lg, para, ar...
7	8	negative	[krn, sebagian, besar, rakyat, indonesia, itu,...
8	9	negative	[ayu, janda, bego, pny, suami, kpn, nikah, lag...
9	10	negative	[anyiennnnggg, suaranya, ancur, banget, lebih,...

Gambar 4. Hasil dataset yang telah di *Tokenizing*.

3. Proses *Stopword Removal* : proses *Stopword Removal* bertujuan untuk menghilangkan atau menghapus kata-kata yang tidak bermanfaat atau kata-kata hubung contoh nya “yang “ atau “ke”. Kata-kata tersebut tidak memberikan kontribusi signifikan dalam membedakan kategori atau topik teks yang berbeda.
4. Proses *Stemming* : proses *stemming* dalam bahasa inggris digunakan untuk mengubah suatu kata menjadi kata dasarnya. Sedangkan proses *stemming* dalam bahasa Indonesia digunakan untuk menghilangkan kata himbuan, contohnya “mencuci” menjadi cuci. Dan proses *stemming* juga bertujuan untuk mengurangi variasi kata sehingga kata-kata yang memiliki akar kata yang sama akan diperlakukan sebagai satu kesatuan, sehingga membantu algoritma klasifikasi dalam memahami pola dan makna teks. Proses *stemming* memerlukan waktu untuk mengubah dan menghilangkan kata himbuan tersebut, semakin banyak kata yang diubah maka semakin lama juga proses yang dijalankan.

#### 2.4. Ekstraksi Fitur (TF-IDF)

TF-IDF (*Term Frequency-Inverse Document Frequency*) metode yang umum digunakan dalam pemrosesan teks untuk mengukur kepentingan relatif suatu kata dalam sebuah dokumen atau koleksi dokumen. Dan metode TF-IDF ini digunakan untuk memboboti suatu kata, semakin banyak kata didalam suatu dokumen maka semakin tinggi nilai bobotnya. Misalkan di dalam suatu dokumen memiliki kata yang berulang, contohnya kata uang muncul berulang kali maka kata uang tersebut memiliki bobot yang tinggi dibandingkan kata yang lain [6].

#### 2.5. Splitting data training dan testing

Splitting data bertujuan untuk memisahkan antara data training dan data testing, data training tersebut akan dibagi menjadi 80 dan 20, yang dimana data training memiliki data 80% sedangkan data testing memiliki 20% data. dari keseluruhan data yang ada di dataset. Pengambilan pembagian data tersebut akan dirandom atau diacak urutan pengambilannya agar data training dan data testing memiliki jenis kategori yang berbeda-beda dan hasil data training tersebut akan menjadi bagus.

#### 2.6. Model Klasifikasi (*Naïve Bayes*)

Pada penelitian ini akan digunakan metode atau algoritma *Naïve Bayes*, Metode atau algoritma *Naive Bayes* adalah algoritma klasifikasi yang didasarkan pada teorema Bayes

dengan asumsi naif bahwa setiap fitur/atribut dalam data adalah independen secara kondisional terhadap kelas target. Pada dasarnya, metode ini menghitung probabilitas posterior kelas target berdasarkan probabilitas prior kelas dan probabilitas fitur. Dan terkait untuk kasus yang dilakukan pada penelitian ini adalah sentiment analisis, akan digunakan algoritma *Naïve Bayes Multinomial* karena algoritma *Naïve Bayes Multinomial* adalah variasi dari metode *Naïve Bayes* yang umum digunakan untuk klasifikasi teks yang mengandung fitur-fitur diskrit atau terhitung. Metode tersebut juga sering digunakan dalam pemrosesan teks, contohnya seperti analisis sentimen, klasifikasi dokumen dan spam filtering [7].

### 3. Hasil dan Pembahasan

Untuk melihat hasil dari performa algoritma yang telah dibuat menggunakan algoritma *Naïve Bayes* perlu dilakukan tahapan evaluasi model. Evaluasi model bertujuan untuk memastikan bahwa model dapat bekerja dengan baik. Evaluasi model dilakukan dengan menggunakan data testing dan metrik evaluasi, Dan metrik evaluasi tersebut akan dilakukan analisis berdasarkan perhitungan Akurasi(*Accuracy*), Presisi(*Precision*), *Recall* dan *F1-score* [6]. *Accuracy* dan *F1-Score* diperoleh dari confusion matrix yang terdiri dari True positive (TP), True negative (TN), False positive (FP), False negative (FN). Rumus yang digunakan untuk evaluasi berbasis nilai dalam confusion matrix dihitung menggunakan persamaan berikut :

$$accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (1)$$

$$precision = \frac{\sum_{l=1}^L TP_l}{\sum_{l=1}^L TP_l + FP_l} \quad (2)$$

$$recall = \frac{\sum_{l=1}^L TP_l}{\sum_{l=1}^L TP_l + FN_l} \quad (3)$$

$$F1 - Score = \frac{\sum_{l=1}^L 2TP_l}{\sum_{l=1}^L 2TP_l + FP_l + FN_l} \quad (4)$$

Adapun Hasil dari performa algoritma *Naïve Bayes* analisis yang telah dilakukan menggunakan data training sebanyak 80% dan data testing 20% dari dataset pada komentar Instagram menghasilkan nilai akurasi 91%, persisi 97%, recall 86%, dan F-1 Score 91%. Dari tingkat akurasi yang didapatkan metode *Naïve Bayes* memiliki nilai hasil yang tinggi. Karena metode *Naïve Bayes* dalam melakukan pengujian klasifikasi, memiliki tingkat ketepatan yang tinggi meskipun sangat sederhana dan efisien.

Table 1 Performa Algoritma Naïve Bayes

Accuracy	Precision	Recall	F-1 Score
91%	97%	86%	91%

Table 2 Performa Sentimen atau kategori

Sentimen	Precision	Recall	F-1 Score
negative	97%	86%	92%
Positive	85%	97%	91%

#### 4. Kesimpulan

Metode *Naïve Bayes* dapat digunakan untuk menganalisis sentiment *cyberbullying*, hasil pengujian pada sistem yang dibuat memperlihatkan bahwa metode *naïve bayes* memberikan nilai akurasi 91%, persisi 97%, recall 86%, dan F-1 Score 91%. Dengan menggunakan splitting atau pembagian data 80/20, 80% data training dan 20% data testing. Dan hasil dari performa masing-masing sentiment atau kategori memiliki nilai yang berbeda yaitu, sentiment negative memiliki nilai persisi 97%, recall 86%, dan F-1 Score 92%. Sedangkan sentiment positive memiliki nilai persisi 85%, recall 97%, dan F-1 Score 91%. Artinya dataset komentar Instagram yang digunakan pada penelitian ini memiliki nilai sentiment negative yang tinggi. Hasil yang diperoleh pada penelitian ini belum cukup akurat, karena pada saat tahapan preprocessing data belum dilakukannya pembersihan data secara bersih, jadi intinya akurasi suatu algoritma atau mesin itu tergantung dari preprocessing datanya, kalau semakin bersih data tersebut di preprocessing maka data tersebut semakin bagus tingkat akurasi nya. Dan juga semakin banyak dataset yang dimiliki akan semakin bagus juga untuk model sistem yang dilatih.

#### Daftar Pustaka

- [1] "Profil Pengguna Internet Indonesia," *Asosiasi Penyelenggara Jasa Internet Indonesia APJII*, 2023. <https://apjii.or.id/>
- [2] "Social media users in Indonesia," *Napoleoncat*, 2023. <https://napoleoncat.com/stats/social-media-users-in-indonesia/2023/>
- [3] A. Akhter, K. A. Uzzal, and M. M. A. Polash, "Cyber Bullying Detection and Classification using Multinomial Naïve Bayes and Fuzzy Logic," *Int. J. Math. Sci. Comput.*, vol. 5, no. 4, pp. 1–12, 2019, doi: 10.5815/ijmsc.2019.04.01.
- [4] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4704–4713, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [5] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," *PLoS One*, vol. 15, no. 5, pp. 1–22, 2020, doi: 10.1371/journal.pone.0232525.
- [6] M. F. Luthfi and K. M. Lhaksamana, "Implementation of TF-IDF Method and Support Vector Machine Algorithm for Job Applicants Text Classification," *J. Media Inform. Budidarma*, vol. 4, pp. 1181–1186, 2020, doi: 10.30865/mib.v4i4.2276.
- [7] S. Ruan, H. Li, C. Li, and K. Song, "Class-specific deep feature weighting for naïve bayes text classifiers," *IEEE Access*, vol. 8, pp. 20151–20159, 2020, doi: 10.1109/ACCESS.2020.2968984.