

ANALISIS SENTIMEN TERHADAP KOMENTAR NEGATIF DI MEDIA SOSIAL FACEBOOK DENGAN METODE KLASIFIKASI *NAÏVE BAYES*

Zaenal¹⁾, Yulita Salim²⁾, Lutfi Budi Ilmawan³⁾

^{1,2,3} Fakultas Ilmu Komputer, Universitas Muslim Indonesia Makassar
^{1,2,3} Jl. Urip Sumoharjo Km.5, Makassar, Sulawesi Selatan

¹ zaenal.official.1@gmail.com; ² yulita.salim@umi.ac.id; ² luthfy13@gmail.com

INFORMASI ARTIKEL	ABSTRAK
Diterima : xx – xx – 20xx Direvisi : xx – xx – 20xx Diterbitkan : xx – xx – 20xx	<p>Facebook merupakan sosial media di Indonesia dengan jumlah akun aktif tertinggi dan paling sering di kunjungi, Media sosial menjadi sarana yang sangat mudah dan bebas untuk beropini, juga memiliki banyak manfaat seperti menuangkan pemikiran dengan membuat status yang didapat dibaca oleh seluruh pengguna media sosial maupun berkomentar mengenai isu-isu terkini, namun dibalik itu semua muncul masalah baru yaitu komentar negatif, salah satunya adalah cyberbullying yang memiliki dampak mendalam dan tahan lama pada korban. Beberapa penelitian melaporkan bahwa korban cyberbullying cenderung mengalami masalah kesehatan mental yang lebih luas, penyalahgunaan narkoba dan ide bunuh diri. Tujuan penelitian ini adalah mengolah data komentar yang diambil dari media sosial Facebook menggunakan pre-processing data, membangun aplikasi prototype filter komentar untuk mengklasifikasikan komentar negatif cyberbullying, dan menguji metode klasifikasi naïve bayes, Data komentar yang digunakan yaitu 300 data training, dan 100 data testing. Setelah melakukan penelitian ini didapatkan bahwa dengan menggunakan pre-processing data mampu menghilangkan karakter atau kata yang tidak dibutuhkan dari komentar, Aplikasi prototype filter komentar yang dibangun telah mampu menyaring komentar cyberbullying, dan hasil pengujian metode klasifikasi naïve bayes menggunakan metode confusion matrix dengan jumlah 100 komentar data testing didapatkan akurasi sebesar 86%, presisi sebesar 84,6153841538461%, recall sebesar 88%, dan f1-score sebesar 86,27450980392156%.</p>
Kata Kunci: Analisis Sentimen Facebook cyberbullying Naïve Bayes	

I. Pendahuluan

Hasil riset *We Are Social Hotsuite* yang dirilis Januari 2019, pengguna media sosial di Indonesia mencapai 150 juta atau sebesar 56% dari total populasi. Jumlah tersebut naik 20% dari survei sebelumnya. Di Indonesia, Facebook merupakan sosial media dengan jumlah akun aktif tertinggi dan paling sering di kunjungi (Aprianti dkk, 2020). Media sosial menjadi sarana yang sangat mudah dan bebas untuk beropini, juga memiliki banyak manfaat seperti menuangkan pemikiran dengan membuat status yang didapat dibaca oleh seluruh pengguna media sosial maupun berkomentar mengenai isu-isu terkini, kebebasan berpendapat telah lama diatur dalam perundang-undangan baik yang tertuang pada hukum internasional Pasal 29 Deklarasi Universal Hak-Hak Asasi Manusia maupun Undang-undang Dasar 1945 pasal 28. Kebebasan mengeluarkan pendapat ini merupakan hak asasi manusia yang paling mendasar. Hak berpendapat mencakup kebebasan berpendapat secara lisan maupun tulisan (Herawati, 2016), namun dibalik itu semua muncul masalah baru yaitu komentar negatif, salah satunya adalah *cyberbullying*.

Cyberbullying dikenal sebagai bentuk “ancaman” atau “serangan” yang dilakukan seseorang terhadap orang lain yang disampaikan melalui pesan elektronik lewat media. *Bullying* terjadi dalam berbagai bentuk diantaranya *bullying* berupa penghinaan, menyebarkan fitnah, mengucilkan orang lain, memperpermalukan, mengintimidasi, serta mengungkapkan rasa kebencian kepada orang lain menggunakan kata kasar. *Cyberbullying* dianggap lebih mudah dilakukan daripada kekerasan konvensional, karena pelaku tidak perlu bertatap muka pada target (Syarif, dkk 2019).

Cyberbullying memiliki dampak yang mendalam dan tahan lama pada korban. Beberapa penelitian melaporkan bahwa korban *cyberbullying* cenderung mengalami masalah kesehatan mental yang lebih luas, penyalahgunaan narkoba dan ide bunuh diri (Goebert dkk, 2011; Gradinger dkk, 2011). Bauman (2019) menemukan bahwa dari 221 remaja partisipan yang pernah mengalami *cyberbullying* menunjukkan peningkatan tekanan emosional. Penelitian yang dilakukan Gradinger, dkk, (2011) menemukan bahwa para korban *bullying* dan *cyberbullying* menemukan penyesuaian yang buruk, agresi, depresi dan gejala psikosomatik lainnya dibandingkan dengan siswa yang belum mengalami.

Sebelum melakukan klasifikasi komentar negatif *cyberbullying*, data komentar tersebut perlu dilakukan *pre-processing* data. *Pre-processing* merupakan suatu proses pengumpulan data mentah untuk diolah menjadi data yang bermanfaat. Data mentah perlu dilakukan proses *pre-processing* karena masih mengandung data yang *incomplete* (data yang masih memiliki kekurangan pada atributnya), *noisy* (data yang masih mengandung *error* dan *outliers*) dan *inconsistent* (data yang tidak konsisten terhadap penamaan, singkatan, *code*) (Syarif dkk, 2019).

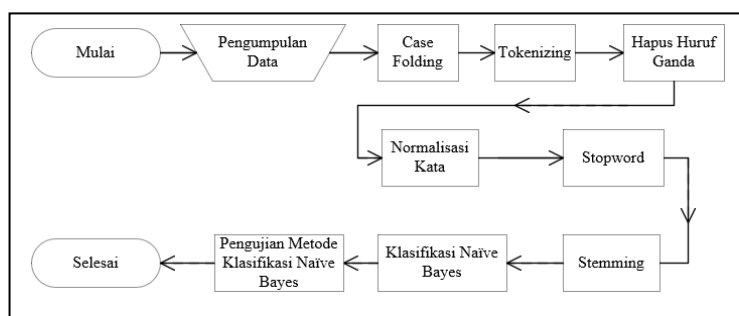
Metode klasifikasi sentimen seperti *Naïve Bayes* (Kang dkk, 2012), *Artificial Neural Network* (Moraes dkk, 2013) (Zhu dkk, 2010), *Support Vector Machine* (Moraes dkk., 2013) (Tan dkk, 2008) diusulkan oleh banyak peneliti (Koncz dkk, 2011) untuk analisis sentimen review restaurant (Kang dkk, 2012), dokumen (Moraes dkk, 2013) (Tan dkk, 2008), dan teks (Zhu dkk, 2010). *Artificial Neural Network* mempunyai kelebihan dalam hal kemampuan untuk generalisasi, yang bergantung pada seberapa baik *Artificial Neural Network* meminimalkan resiko empiris namun *Artificial Neural Network* mempunyai kelemahan dimana menggunakan data pelatihan cukup besar (Vapnik, 1999). *Support Vector Machine* mempunyai kelebihan yaitu bisa diterapkan untuk data yang berdimensi tinggi, tetapi *Support Vector Machine* sulit untuk digunakan untuk data dengan jumlah yang besar (Nugroho dkk, 2003). *Naïve Bayes* dapat mengatasi kelemahan dari *Artificial Neural Network* dan *Super Vector Machine* yaitu masalah jumlah data yang digunakan. Metode *Naïve Bayes* dapat diterapkan dalam data set yang besar (Sulistiono, 2015) juga *Naïve Bayes* masih mampu bekerja dengan baik dengan ukuran data set yang kecil (Kohavi, 2011). Penelitian yang dilakukan oleh Ma'arif (2016) membandingkan metode *Naïve Bayes* dengan *Super Vector Machine* didapatkan hasil model klasifikasi yang dihasilkan oleh *Naïve Bayes* memiliki performa yang lebih tinggi dibandingkan dengan model yang dihasilkan dari metode *Super Vector Machine*, pada penelitian lain yang dilakukan Sulistiono (2015) membandingkan metode *c4.5*, *Neural Network*, dan *Naïve Bayes* didapatkan hasil metode *Naïve Bayes* memiliki tingkat akurasi yang paling tinggi.

Naïve Bayes yang merupakan teknik pembelajaran mesin yang berbasis probabilitas. *Naïve Bayes* adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks (Routray, 2013). Pada penelitian yang dilakukan oleh Domingos dan Pazzani (1997) *Naïve Bayes* dapat bekerja dengan baik bahkan dengan adanya kehadiran dari fitur yang memiliki dependensi yang kuat pada dataset.

Tujuan penelitian yang ingin dicapai yaitu mengolah data komentar negatif *cyberbullying* menggunakan *pre-processing* untuk menghilangkan kata atau karakter yang tidak dibutuhkan, mengukur tingkat akurasi, presisi, *recall*, dan *F1-score* metode klasifikasi *naïve bayes* dalam mengklasifikasikan komentar negatif *cyberbullying*, dan membangun prototype filter komentar negatif *cyberbullying*, dimana komentar yang ditulis dan terdeteksi sebagai komentar *cyberbullying*, maka user tidak bisa memposting komentar dan sistem memberikan peringatan terhadap komentar tersebut

II. Metode

Pada penelitian ini, tahapan-tahapan dalam melakukan analisis sentiment terhadap komentar negatif dimedia sosial Facebook terlihat pada gambar *flowchart* sebagai berikut:



Gambar 1. *Flowchart* tahapan penelitian

A. Pengumpulan Data

Komentar yang digunakan bersumber dari akun Facebook presiden Jokowi, Rocky Gerung, iNews, Syahrini, Kekey Puteri Cantika, Barbie Kumalasari, dan Farhat Abbas, untuk proses pengumpulan data komentar dilakukan secara manual. Jumlah komentar yang digunakan sebesar 400 komentar yang terdiri dari 300 komentar *data training*, dan 100 komentar untuk *data testing*. Komentar yang telah dikumpulkan diberi kelas secara manual yaitu kelas *cyberbullying* atau kelas *non cyberbullying*, untuk kategori komentar yang diberi kelas *cyberbullying* yaitu *Called Name* (Pemberian Nama Negatif), *Threatened Physical Harm* (Mengancam Keselamatan Fisik), dan *Opinion Slammed* (Pendapat Yang Merendahkan), selain dari pada itu komentar diberi kelas *non cyberbullying*.

B. Case Folding

Case folding adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil contohnya kalimat “Sehat selalu pak Presiden” menjadi “sehat selalu pak presiden”. Hanya huruf “a” sampai dengan “z” yang diterima. Karakter selain huruf dihilangkan dan dianggap *delimiter* (melebihi batas).

C. Tokenizing

Tokenizing atau *parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya, contohnya:

Tabel 1. *Tokenizing*

Kalimat	Hasil Tokenizing			
sehat selalu pak presiden	sehat	selalu	pak	presiden

D. Hapus Huruf Ganda

Menghilangkan huruf ganda merupakan tahap untuk menormalkan huruf berulang seperti kata “seeenaaang” menjadi kata “senang”

E. Normalisasi Kata

Normalisasi kata merupakan tahap merubah kata prokem, kata tidak baku, ataupun kata singkatan menjadi kata baku. Contohnya kata prokem “gua” menjadi “saya”, kata singkatan “yg” menjadi “yang”.

F. Stopword

Stopword elimination merupakan proses penghilangan kata yang termasuk di dalam daftar kata tidak penting (*stoplist*). Jika termasuk di dalam *stoplist* maka kata tersebut akan dihilangkan, contohnya yaitu: “dan, maka, di, ke, walaupun, yang, ini, itu, disini”. Untuk daftar *stopword* atau *stoplist* yang digunakan, bersumber dari penelitian Tala (2003).

G. Stemming

Stemming merupakan sebuah proses ekstraksi dari sebuah kata yang digunakan untuk mendapatkan kata dasar, misalnya kata “menghasilkan” menjadi kata “hasil”.

H. Klasifikasi Naïve Bayes

Metode klasifikasi *Naïve Bayes* termasuk dalam metode *machine learning* yang merupakan salah satu algoritma yang sederhana namun memiliki kemampuan dan akurasi yang tinggi. Metode klasifikasi *Naïve Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* dengan asumsi independensi yang kuat. Jadi dalam *Naïve Bayes*, model yang digunakan adalah model fitur independen. Maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Berikut adalah rumus dari metode klasifikasi *Naïve Bayes* untuk menghitung nilai probabilitas setiap kata:

1. Klafikasi Naïve Bayes

Untuk melakukan klasifikasi *naïve bayes* menggunakan rumus (1)

$$CNB = \arg \max c P(c) \prod_{i=1}^n P(F_i|c) \quad (1)$$

Keterangan

$\arg \max$: nilai *posterior probability* terbesar suatu kelas

$P(F_i|c)$: nilai *coditional probability*

$P(c)$: nilai *prior probability*

2. Hitung nilai *prior probability*

Untuk menghitung *prior probability* menggunakan rumus (2)

$$P(c) = \frac{Nc}{Ndoc} \quad (2)$$

Keterangan

$P(c)$: Nilai probabilitas data latih

Nc : jumlah data *cyberbullyng* atau *non cyberbullying*

$Ndoc$: Total data latih

3. Hitung *conditional probability*

Untuk mengetahui *conditional probability* yang mengasumsikan kemunculan kata pada setiap dokumen dapat dinyatakan dengan persamaan (3)

$$P(f_i|c) = \frac{\text{count}(f_i, c)}{\sum_{i=1}^n f \in v \text{count}(f_i, c)} \quad (3)$$

Untuk menghindari terjadinya nilai nol (0) pada probabilitas setiap kata maka dapat dinyatakan dengan persamaan (4).

$$P(f_i|c) = \frac{\text{count}(f_i, c) + 1}{(\sum_{i=1}^n f \in v \text{count}(f, c)) + V} \quad (4)$$

Keterangan

$P(f_i c)$:	nilai probabilitas setiap kata
f_i, c	:	jumlah kata tertentu yang terdapat pada seluruh dokumen pada kelas c
c	:	kelas (<i>cyberbullying</i> atau <i>non cyberbullying</i>)
F, c	:	jumlah seluruh kata yang terdapat pada kelas c .
V	:	Jumlah kata berbeda pada semua dokumen

I. Pengujian Metode Klasifikasi *Naïve Bayes*

Metode pengujian *confusion matrix* merupakan salah satu metode yang digunakan untuk menguji akurasi, presisi, *recall*, dan *F1-score* dari sebuah metode klasifikasi. Berikut adalah penjelasan tentang akurasi, presisi, *recall*, dan *F1-score*:

1. Akurasi

Akurasi adalah tingkat kedekatan nilai prediksi dengan nilai actual, untuk menghitung akurasi menggunakan rumus (5)

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

2. Presisi

Presisi adalah tingkat ketepatan atau ketelitian dalam pengklasifikasian untuk menghitung presisi menggunakan rumus (6)

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

3. Recall

Recall berfungsi untuk mengukur proporsi positif aktual yang benar diidentifikasi, untuk menghitung *recall* menggunakan rumus (7)

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \quad (7)$$

4. F1-score

digunakan untuk mengetahui keseimbangan antara presisi dan *recall* yang didapat dari sistem yang akan dibangun. untuk menghitung *f1-score* menggunakan rumus (8)

$$F1\text{-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (8)$$

III. Hasil dan Pembahasan

A. Preprocessing

Komentar yang akan dilakukan proses klasifikasi terlebih dahulu harus melewati prose *pre-processing* karena masih mengandung data yang *incomplete* (data yang masih memiliki kekurangan pada atributnya),

noisy (data yang masih mengandung *error* dan *outliers*) dan *inconsistent* (data yang tidak konsisten terhadap penamaan, singkatan, kode). Hasil *pre-processing* komentar terlihat pada gambar 2.

Analisis Sentimen															
<p>Form Input Komentar</p> <p>Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡</p> <p>Posting Komentar</p> <p>Tampilkan lebih sedikit</p> <p>Unknown diharapkan selalu menggunakan helm untuk menjaga keselamatan diri saat bermotor</p> <p>Unknown diharapkan selalu menggunakan helm untuk menjaga keselamatan diri saat bermotor</p> <p>Unknown diharapkan selalu menggunakan helm untuk menjaga keselamatan diri saat bermotor</p>	<p>Tabel Pre-processing</p> <table> <tr> <th>prerprocessing</th><th>Hasil Preprocessing</th></tr> <tr> <td>Case Folding</td><td>maunya kau yg kena korona. karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡</td></tr> <tr> <td>Tokenizing</td><td>maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah</td></tr> <tr> <td>Hapus Huruf Ganda</td><td>maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah</td></tr> <tr> <td>Normalisasi</td><td>maunya kau yang kena korona karena hidupmu hanya buat susah orang dasar manusia sampah</td></tr> <tr> <td>Stopword</td><td>maunya kau kena korona hidupmu susah orang dasar manusia sampah</td></tr> <tr> <td>Stemming</td><td>mau kau kena korona hidup susah orang dasar manusia sampah</td></tr> </table>	prerprocessing	Hasil Preprocessing	Case Folding	maunya kau yg kena korona. karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	Tokenizing	maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah	Hapus Huruf Ganda	maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah	Normalisasi	maunya kau yang kena korona karena hidupmu hanya buat susah orang dasar manusia sampah	Stopword	maunya kau kena korona hidupmu susah orang dasar manusia sampah	Stemming	mau kau kena korona hidup susah orang dasar manusia sampah
prerprocessing	Hasil Preprocessing														
Case Folding	maunya kau yg kena korona. karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡														
Tokenizing	maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah														
Hapus Huruf Ganda	maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah														
Normalisasi	maunya kau yang kena korona karena hidupmu hanya buat susah orang dasar manusia sampah														
Stopword	maunya kau kena korona hidupmu susah orang dasar manusia sampah														
Stemming	mau kau kena korona hidup susah orang dasar manusia sampah														

Gambar 2. Hasil *pre-processing*

B. Aplikasi *Prototype Filter* Komentar

Aplikasi *Prototype Filter* Komentar merupakan aplikasi analisis sentiment terhadap komentar negatif yang termasuk sebagai *cyberbullying*. Komentar yang akan ditulis oleh *user* jika terdeteksi sebagai komentar *cyberbullying*, sistem akan menampilkan peringatan bahwa komentar yang ditulis merupakan komentar *cyberbullying*, dan komentar tersebut tidak dapat di-*posting*, sebagai contoh komentar “Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡” ditulis pada *form input* komentar, ketika akan di-*posting* menampilkan peringatan bahwa komentar yang ditulis merupakan komentar *cyberbullying*, dan komentar tersebut tidak akan ter-*posting* seperti yang terlihat pada gambar 3.

Analisis Sentimen

Home Pengujian

Form Input Komentar

Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡

Posting Komentar

Tampilkan semua komentar

Unknown

diharapkan selalu menggunakan helm untuk menjaga keselamatan diri saat bermotor

Unknown

diharapkan selalu menggunakan helm untuk menjaga keselamatan diri saat bermotor

Unknown

diharapkan selalu menggunakan helm untuk menjaga keselamatan diri saat bermotor

Unknown

diharapkan selalu menggunakan helm untuk menjaga keselamatan diri saat bermotor

Tabel Pre-processing

prerprocessing	Hasil Preprocessing
Case Folding	maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah
Tokenizing	maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah
Hapus Huruf Ganda	maunya kau yg kena korona karna hidupmu cuma bikin susah org dasar manusia sampah
Normalisasi	maunya kau yang kena korona karena hidupmu hanya buat susah orang dasar manusia sampah
Stopword	maunya kau kena korona hidupmu susah orang dasar manusia sampah
Stemming	mau kau kena korona hidup susah orang dasar manusia sampah

Tabel Analisis

Tabel Komentar tersaring

No	Komentar	Jenis Komentar
1	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying
2	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying
3	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying
4	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying
5	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying
6	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying
7	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying
8	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying
9	Maunya kau yg kena korona. Karna hidupmu cuma bikin susah org... dasar manusia sampah 😡😡😡	cyberbullying

Oops...

Terdeksi Sebagai Komentar Cyberbullying!

OK

Gambar 3. Menulis Komentar

C. Pengujian Confusion Matrix

Pada tahapan ini dilakukan pengujian metode klasifikasi *naïve bayes* untuk mengukur akurasi, presisi, *recall*, dan *f1-score*. Jumlah komentar yang dijadikan *data testing* yaitu 100 data yang terbagi menjadi 50 komentar *cyberbullying* dan 50 komentar *non cyberbullying* yang diberi kelas secara manual oleh penulis.

dimasukkan kedalam tabel confusion matrix yang selanjutnya akan dilakukan perhitungan akurasi, presisi, *recall*, dan *f1-score*. Hasil pengujian dapat dilihat pada gambar 4.

Mulai Pengujian			
Tabel Data Testing			Tabel Confusion Matrix
Show 10 entries	Search:		
No	Komentar	Kelas	
1	Maju amburadunya..maju ngibulnya. maju asbunya.. lanjutkan kedunguanmu bos	cyberbullying	
2	Pasukan sok mau nasehati orang. dirinya aja sok bijak. bangsat sialan	cyberbullying	
3	Itu salah satu prestasi sibotak sampah yg gk bcus ngurus. Anaknya	cyberbullying	
4	Yg bilang belum muhrim. aduuuhh sok suci kali anda. Pentil tetek lu udah berapa lidah yg singgah	cyberbullying	
5	Fotonya kya patung2 ala2 romawi yg gak pake celana dan TT nya kemana2	cyberbullying	
6	Kepada pak joko tolong para penjiatmu dikendalikan dungunya	cyberbullying	
7	Nyengir,cebong mau lompat ke kolam. biar dungunya penuh sekolah,hahahah	cyberbullying	

#	Positif	Negative
Positif	44 (True Positif)	8 (False Positif)
Negative	6 (False Negative)	42 (True Negative)

Hasil Pengujian

Akurasi : $(TP + TN) / (TP + TN + FP + FN)$
: $(44 + 42) / (44 + 42 + 8 + 6)$
: 86 %

Presisi : $TP / (TP + FP)$
: $44 / (44 + 8)$
: 84.61538461538461 %

Recall : $TP / (TP + FN)$
: $44 / (44 + 6)$
: 88 %

F1-score
: $2 * ((Presisi * Recall) / (Presisi + Recall))$
: $2 * (84.61538461538461 * 88) / (84.61538461538461 + 88)$
: 86.27450980392156 %

Gambar 4. Pengujian *confusion matrix*

Pada gambar 4 Setelah melakukan pengujian didapatkan nilai *true positif* dengan nilai 44 yang berarti berhasil mengklasifikasikan 44 komentar *cyberbullying* dengan benar, *false negative* dengan nilai 6 yang berarti terdapat 6 komentar *cyberbullying* yang diklafikasikan sebagai komentar *non cyberbullying*, *true negative* dengan nilai 42 yang berarti ada 42 komentar *non cyberbullying* yang diklasifikasikan dengan benar, *false positif* dengan nilai 8 yang berarti ada 8 komentar *non cyberbullying* yang diklasifikasikan sebagai komentar *cyberbullying*. Nilai *true positif*, *true negative*, *false positif*, dan *false negative* yang digunakan untuk perhitungan akurasi, presisi, *recall*, dan *f1-score*, untuk perhitungan akurasi menggunakan persamaan (5)

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

$$\text{Akurasi} = \frac{44+42}{44+42+8+6} \times 100\%$$

$$= 86 \%$$

Maka didapatkan nilai akurasi sebesar 86% seperti yang terlihat pada gambar 4 pada hasil pengujian, untuk perhitungan presisi menggunakan persamaan (6).

$$\text{Persisi} = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$\text{Persisi} = \frac{44}{44+8} \times 100\%$$

$$= 84.61538461538461 \%$$

Maka didapatkan nilai presisi sebesar 84.61538461538461 % seperti yang terlihat pada gambar 4 pada hasil pengujian, kemudian untuk perhitungan *recall* menggunakan persamaan (7)

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

$$\text{Recall} = \frac{44}{44+6} \times 100\%$$

$$= 88 \%$$

Maka didapatkan nilai *recall* sebesar 88% seperti yang terlihat pada gambar 5.7 pada hasil pengujian, dan perhitungan *f1-score* menggunakan rumus (8)

$$F1\text{-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (8)$$

$$F1\text{-score} = (2 \times \frac{84.61538461538461 \% \times 88 \%}{84.61538461538461 \% + 88 \%}) \times 100 \%$$

$$= 86.27450980392156 \%$$

Maka didapatkan nilai *f1-score* sebesar 86.27450980392156 % seperti yang terlihat pada gambar 4 pada hasil pengujian.

IV. Kesimpulan

Berdasarkan seluruh hasil penelitian analisis sentiment terhadap komentar negatif di media sosial Facebook menggunakan metode klasifikasi *naïve bayes* dapat disimpulkan sebagai berikut:

- A. Dengan menggunakan *pre-processing* data mampu menghilangkan karakter atau kata yang tidak dibutuhkan seperti karakter tanda baca, mengubah kata prokem menjadi kata baku, menghilangkan kata umum dari komentar yang digunakan sebagai data analisis sentimen.
- B. Aplikasi *prototype filter* komentar telah mampu menyaring komentar yang terdeteksi sebagai komentar *cyberbullying* dan sistem memberikan peringatan kepada *user* jika komentar yang tulis terdeteksi sebagai komentar *cyberbullying* dan komentar tersebut tidak dapat di-*posting* oleh *user*.
- C. Berdasarkan Hasil pengujian terhadap metode klasifikasi *naïve bayes* yang dijelaskan pada sub bab pengujian sistem dengan menggunakan 100 komentar sebagai data uji yang terdiri dari 50 komentar *cyberbullying* dan 50 komentar *non cyberbullying*, mampu mengklasifikasikan 44 komentar *cyberbullying* dengan benar, 6 komentar *cyberbullying* diklasifikasikan sebagai *non cyberbullying*, mengklasifikasi 42 *non cyberbullying* dengan benar, 8 komentar *non cyberbullying* diklasifikasikan sebagai *cyberbullying*. Hasil klasifikasi *data testing* dilakukan perhitungan akurasi, presisi, *recall*, dan *f1-score* maka diperoleh akurasi yaitu 86%, presisi yaitu 84,6153841538461%, *recall* yaitu 88%, dan *f1-score* yaitu 86,27450980392156%.

Ucapan Terima Kasih

Dengan bantuan berbagai pihak dalam menyelesaikan paper ini, sehingga sudah sepantasnya penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. H. Basri Modding, SE., M.Si selaku Rektor Universitas Musli Indonesia, beserta para Wakil Rektor.iv
2. Bapak Purnawansyah, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Muslim Indonesia, beserta para Wakil Dekan.
3. Bapak Tasrif Hasanuddin, S.T., M.Cs selaku Ketua Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Muslim Indonesia.
4. Ibu Yulita Salim, S.Kom., M.T selaku pembimbing utama yang telah banyak membantu dan membimbing dalam penyelesaian tugas akhir ini.
5. Bapak Lutfi Budi Ilmawan, S.Kom., M.Cs selaku pembimbing pemdamping yang telah banyak membantu dan membimbing dalam penyelesaian tugas akhir ini.
6. Penguji pertama bapak Huzain Azis, S.Kom., M.Cs yang telah memberikan kritik dan saran dalam menyelesaikan tugas akhir ini.
7. Penguji kedua bapak Syahrul Mubarak Abdullah, S.Kom., M.Kom yang telah memberikan kritik dan saran dalam menyelesaikan tugas akhir ini.
8. Penguji ketiga ibu Herdianti, S.Si., M.Eng yang telah memberikan kritik dan saran dalam menyelesaikan tugas akhir ini.
9. Seluruh dosen pengajar jurusan Teknik Informatika Universitas Muslim Indonesia Makassar, yang telah mendidik dan memberikan berbagai bekal pengetahuan yang tak ternilai harganya kepada penulis selama mengikuti perkuliahan.
10. Kepada seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah dengan tulus ikhlas memberikan doa dan motivasi kepada penulis sehingga dapat menyelesaikan tugas akhir.

Daftar Pustaka

- [1] Akbar, Muhammad Alam., Utari, Prahastiwi. *Cyberbullying Pada Media Sosial (Studi Analisis Isi tentang Cyberbullying pada Remaja di Facebook)*. Program Studi Ilmu Komunikasi Fakultas Ilmu Sosial dan Ilmu Politik Universitas Sebelas Maret Surakarta. 1-20.
- [2] Aprianti, Nova., Nasution, Itto Nesyia., Aiyuda, Nurul. 2020. Fungsi Eksekutif pada Prasangka Pengguna Facebook terhadap Presiden RI. *Psychopolytan: Jurnal Psikologi*, volume 3, pp. 63-70.
- [3] Bauman, S. (2009). *Cyberbullying in a rural intermediate school: An exploratory study. The Journal of Early Adolescence, online first*, 1-31.
- [4] Domingos, P., Pazzani, M. 1997, *On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, Machine Learning*, 29, 103-130
- [5] Goebert, D., Else, I., Matsu, C., Chung-Do, J., & Chang, J.Y. (2011). *The impact of cyberbullying on substance use and mental health in a multiethnic sample. Maternal Child Health Journal*, 15, 1282-1286.
- [6] Gradinger, P., Strohmeier, D., & Spiel, C. (2009). *Traditional bullying, and cyberbullying: identification*

- of risk groups for adjustment problems. *Journal of Psychology*, 217(4), 205-213.
- [7] Harijati, Servasius Dwi. 2019. Analisis Sentimen Pada Twitter Menggunakan *Multinomial Naive Bayes*. Universitas Sanata Dharma Yogyakarta.
- [8] Hasibuan, Zainal A. 2007. Metodologi Penelitian Pada Bidang Ilmu Komputer dan Teknologi Informasi. Fakultas Ilmu Komputer Universitas Indonesia.
- [9] Herawati, Dewi Maria. 2016. Penyebaran Hoax dan *Hate Speech* sebagai Representasi Kebebasan Berpendapat. *Promedia*, Volume 2, pp. 138-155
- [10] Ismiati, Maria, Bellanir. 2018. Deteksi Komentar Negatif Di Instagram Menggunakan Algoritma *Naive Bayes Classifier*. Prosiding, Fakultas Teknik Universitas Wahid Hasyim.
- [11] Ilmawan, Lutfi Budi. 2014. Aplikasi Mobile Untuk Analisis Sentimen Pada Google Play.
- [12] Kang, H., Yoo, S. J., & Han, D. 2012. *Senti lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews*. *Expert Systems with Applications*, Volume 39, pp. 6000–6010.
- [13] Kohavi, Ron. 2011. *Scaling Up the Accuracy of Naive-Bayes Classifiers : a Decision-Tree Hybrid*. In *KDD-96 Proceedings*, AAAI, California.
- [14] Kondz, P., & Paralic, J. 2011. *An approach to feature selection for sentiment analysis*. In 2011 15th IEEE International Conference on Intelligent Engineering Systems (pp. 357–362). IEEE.
- [15] Liantoni, Febri. 2016. “Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor.” *Jurnal ULTIMATICS* 7(2): 98–104.
- [16] Ling, Juen., Kencana, I Putu Eka N., Oka, Tjokorda Bagus. 2014. Analisis Sentimen Menggunakan Metode *Naive Bayes Classifier* Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, volume. 3, pp. 92-99.
- [17] Luqyana, Wanda Athira., Cholissodin, Imam Perdana., Rizal Setya. 2018. Analisis Sentimen *Cyberbullying* pada Komentar Instagram dengan Metode Klasifikasi *Support Vector Machine*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, volume 2, pp. 4704-4713.
- [18] Ma’arif, Muhammad Rifqi. 2016. Perbandingan *Naive Bayes Classifier* Dan *Support Vector Machine* Untuk Klasifikasi Judul Artikel. *JISKA*, volume 1, pp. 90-93.
- [19] Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document Level Sentiment Classification: an Empirical Comparison between SVM and ANN. *Expert Systems with Applications*, Volume 40, pp. 621–633.
- [20] Nugroho, A. S., Witarto, A. B., & Handoko, D. 2003. Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika. *IlmuKomputer*.
- [21] Nurdiana, O., Jumadi., dan Nursantika, D. 2016. Perbandingan Metode Cosine Similarity dengan Metode *Jaccard Similarity* pada Aplikasi Pencarian Terjemahan Al-Qur’an dalam Bahasa Indonesia. *Jurnal Online Informatika* Volume 1, pp, 59-63.
- [22] Pratika, G Dinda. 2015. *Action Research* Metode Penelitian. Akuntansi & Sistem Informasi Bina Nusantara University
- [23] Routray, Preeti. 2013. *A Survey on Sentiment Analysis*. *International Journal of Computer Applications*, volume 2, pp. 1-8.
- [24] Sulistiono, Heru. 2015. “Kajian Penerapan Algoritma C4.5, Neural Network Dan *Naive Bayes* Untuk Klasifikasi Mahasiswa Yang Bermasalah Dalam Registrasi.” *Faktor Exacta* 8(4): 305–15.
- [25] Susyanto, Teguh. 2018. “Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani.” *Jurnal Ilmiah SINUS*.
- [26] Sasongko, Theopilus Bayu. 2016. Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM. *Jurnal Teknik Informatika dan Sistem Informasi*, Volume 2, pp. 244-253.
- [27] Sipayung, Yoannes Romando., Sulistyowati, Reny., Wibisono, Iwan Setiawan. 2019. Identifikasi Komentar Negatif Berbahasa Indonesia Pada Instagram Dengan Metode *K-Means*. *Jurnal Prodi Teknik Informatika UNW “Multimatrix”*, volume 2, pp, 6-8.
- [28] Syarif, Rizky Dhian., Herdiani, Anisa, S.T., M.T., Astuti, Widi, S.T., M.Kom. 2019. Identifikasi *Cyberbullying* pada Komentar Instagram menggunakan Metode *Lexicon-Based* dan *Naive Bayes Classifier* (Studi kasus: Pemilihan Presiden Indonesia Tahun 2019). Prosiding. Universitas Telkom Bandung.
- [29] Tala, Fadillah Z. 2003. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Master of Logic Project Institute for Logic, Language and Computation Universiteit van Amsterdam The Netherlands
- [30] Tan, S., & Zhang, J. 2008. *An empirical study of sentiment analysis for chinese documents*. *Expert Systems with Applications*, Volume 34.
- [31] Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, Volume 10, pp. 988–99.
- [32] Zhu, J., Xu, C., & Wang, H. (2010). *Sentiment classification using the theory of ANNs*. *The Journal of China Universities of Posts and Telecommunications*, Volume 17, pp. 58–62.

