ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

ANALISA SENTIMEN *CYBERBULLYING* DI JEJARING SOSIAL TWITTER DENGAN ALGORITMA NAÏVE BAYES

Fajar Agus Maulana[1], Iin Ernawati[2]
Program Studi Informatika, Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia.
fajaragusmaulana@gmail.com

Abstrak. Tweet atau cuitan yang mengandung unsur cyber bullying bisa menyinggung dan menimbulkan permusuhan antar pengguna twitter. Penelitian ini menggunakan algoritma Naïve Bayes untuk proses klasifikasi dari tweet. Data yang digunakan ialah bersumber dari akun yang sering membuat tweet mengenai politik dengan menggunakan Twitter API (Application Programming Interface). Pada data latih tweet yang tidak mengandung unsur cyber bullying diberikan label positif, sedangkan tweet yang mengandung unsur cyber bullying diberikan label negatif. Hasil pengujian dengan data uji real time pada tanggal 12 Mei 2020 pukul 01.00 WIB mendapatkan nilai akurasi sebesar 76%. Metode ini cukup baik dalam mengklasifikasikan tweet positif dan negatif, namun pada proses pengujian penelitian ini dalam mendeteksi tweet yang mengandung unsur cyber bullying masih kurang baik, dikarenakan masih terdapatnya tweet yang tidak mengandung unsur cyber bullying didalam data latih yang memiliki label tweet negatif.

Kata Kunci: Tweet, Cyberbullying, Klasifikasi Naïve Bayes.

1 Pendahuluan

Media sosial ini semakin memperluas jangkauan sosial setiap penggunanya. Di media sosial, kita dapat menemukan, berkenalan dan berinteraksi dengan orang asing secara mudah. Kemudahan ini membuat setiap orang bebas (hampir tanpa batas) berbagi informasi dan saling berpendapat [1].
Namun saat ini banyak oknum pengguna media sosial yang melakukan penyalahgunaan untuk melakukan penghinaan ataupun pencemaran nama baik, hal tersebut masuk kedalam kategori cyberbullying. Cyberbullying atau intimidasi dunia maya atau penindasan dunia maya diartikan sebagai sebuah serangan yang dilakukan dengan sengaja dan dilakukan dalam dunia maya terhadap seseorang. Intimidasi atau pelecehan secara verbal secara terus menerus yang dilakukan didunia maya. cyberbullying dapat menyebabkan korban yang di intimidasi mengalami gangguan emosional. Berlatarkan kasus tersebut, maka perlu mengambil tindakan pencegahan agar tidak bermunculan korban cyberbullying. Menurut undang undang ITE, cyberbullying merupakan hal yang melanggar undang undang [2].
Salah satunya dengan melakukan analisis sentiment untuk mendeteksi tweet-tweet yang mengandung unsur cyberbullying. Penelitian ini menggunakan metode klasifikasi dengan Algoritma Naïve Bayes untuk mendeteksi tweets yang mengandung unsur cyberbullying di jejaring sosial twitter. Algoritma tersebut digunakan karena merupakan algoritma klasifikasi yang dapat mengolah data dalam jumlah besar dan menghasil akurasi yang cukup tinggi .

2 Landasan Teori

2.1 Analisis Sentimen

Analisis sentimen merupakan sebuah teknik yang digunakan untuk mengidentifikasi teks dapat dikategorikan masuk kedalam sentimen positif maupun sentimen negatif [3]. Dana nalisis sentimen adalah proses untuk menentukan emosi, sikap dan opini yang ada didalam teks dan diklasifikasikan menjadi sentimen positif maupun negatif [4].
Dari pendapat tersebut analisis sentimen merupakan proses untuk menentukan opini atau sentimen yang dikategorikan sebagai sentimen positif maupun sentimen negatif berdasarkan teks yang dibuat oleh seseorang.

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

2.2 Praproses Data

Praproses data dilakukan untuk membersihkan data yang masih belum siap digunakan agar data tersebut dapat diolah pada tahap selanjutnya. Pada tahap ini dilakukan untuk menghilangkan data yang tidak sesuai sehingga menjadi data yang dapat untuk diklasifikasikan. Proses ini sangat penting dalam analisis sentimen, karena social media banyak mengandung kata atau kalimat yang memiliki banyak noise, tidak formal atau menggunakan slang word serta tidak terstruktur.

2.2.1 Transform Case

Transform cases adalah proses dimana semua huruf yang ada pada data diubah sesuai dengan keinginan, seperti mengubah uppercase menjadi lowercase atau sebaliknya Pembersihan Data [2].

2.2.2 Pembersihan Data

Pembersihan data dilakukan untuk menghilangkan beberapa data tidak memiliki value. Data yang harus dihilangkan ialah seperti username yang di-mention, menghapus hashtag, dan menghapus URL.

2.2.3 Stopword Removal

Stopword removal merupakan proses penghilangan kata-kata yang tidak berkontribusi banyak pada isi dokumen [7].

2.2.4 Stemming

Stemming merupakan proses menghilangkan imbuhan dan sufiks pada kata dalam data. Stem (akar kata) adalah menghilangkan imbuhan (awalan atau akhiran) untuk mendapatkan kata inti yang terdapat di dalamnya, seperti kata “menghilang” setelah proses stemming imbuhan meng- menghilang berubah menjadi “hilang”.

2.2.5 Tokenisasi

Tokenisasi adalah sebuah proses untuk memotong teks menjadi bagian yang disebut token, yaitu sebuah instance dari urutan karakter dalam beberapa dokumen tertentu yang dikelompokkan bersama sebagai unit semantik yang berguna untuk diproses. [8]

2.3 Algoritma Naïve Bayes

Naïve bayes. Menghitung probabilitas dengan ketentuan bahwa class keputusan adalah benar karena vektor informasi objek. mengasumsikan bahwa atribut obyek ini independent [9]. Probabilitas yang ada dalam proses untuk menghitung perkiraan akhir sebagai jumlah frekuensi dari master tabel keputusan. Berikut merupakan persamaan teorema naïve bayes classifier:

$$P(c|d) = \frac{P(c)\times P(d|c)}{P(d)} \qquad (1)$$

Keterangan:

c : Hipotesis data merupakan suatu class spesifik

d : Data dengan class yang belum diketahui.

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

P(c|d) : Posterior, probabilitas hipotesis A berdasarkan kondisi W.

P(c) : Prior, probabilitas hipotesis A

P(d|c) : Likelihood, probabilitas W berdasarkan kondisi pada hipotesis A.

P(d) : Evidence, probabilitas W.

Dengan teorema bayes maka penelitian ini akan mengimplementasikan teorema bayes dengan sebagai berikut

$$c_{MAP} = argmax \frac{P(c)\times P(d|c)}{P(d)} \quad (2)$$

Nilai p(d) dapat diabaikan karena nilainya konstan untuk semua c sehingga persamaan (2) dapat ditulis sebagai berikut:

$$c_{MAP} = argmax P(c) \times P(d|c) (3)$$

Dengan pendekatan Naïve Bayes yang mengasumsikan bahwa setiap kata dalam setiap kategori adalah tidak bergantung satu sama lain, maka perhitungan dapat lebih disederhanakan dan dapat dituliskan sebagai berikut [10] :

$$c_{MAP} = \underset{c\in V}{arg\,max}\ P(c) \prod_i P(W_i|c) \quad (4)$$

Untuk menghindari division of zero maka dapat menggunakan metode laplace smoothing untuk menghindari nilai pembaginya 0 [11]. Nilai P(Wi|c) dan P(c) dapat dihitung dengan menggunakan persamaan sebagai berikut:

$$P(W_i|c) = \frac{Count(W_i,c)+1}{|c|+|V|} \quad (5)$$

$$P(\ c) = \frac{|\ doc\ c\ |}{|document|} \quad (6)$$

Keterangan:

P(Wi|c) : Probabilitas kata Wi pada kelas c.

Count(Wi,c) : jumlah kemunculan kata Wi pada kelas c.

|c| : jumlah semua kata pada kelas c.

|V| : Jumlah keseluruhan kata.

$P(\ c)$: Peluang kemunculan suatu dokumen yang memiliki kategori j.

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

$doc\ j$: Jumlah dari dokumen untuk tiap kategori j.

$|document|$: Jumlah dokumen dari setiap kategori.

2.4 Evaluasi

Evaluasi pada penelitian ini menggunakan confusion matrix. Evaluasi dilakukan untuk mengetahui nilai performansi dari sistem yang sudah dibuat berdasarkan hasil dari klasifikasi [12]. Parameternya berdasarkan dari tabel confusion matrix. Tabel dibawah merupakan confusion matrix untuk klasifikasi dua class.

Tabel I. Tabel Confusion Matrix

		True Class	
		Positive	Negative
Predicted class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Keterangan :

True Positive (TP), data yang diprediksi positif dan nyatanya positif

True Negative (TN), data yang diprediksi negatif dan nyatanya negatif.

False Positive (FP), data yang diprediksi positif dan nyatanya negatif.

False Negative (FN), data yang diprediksi negatif dan nyatanya positif.

Keempat parameter tersebut digunakan untuk menghitung

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (7)$$

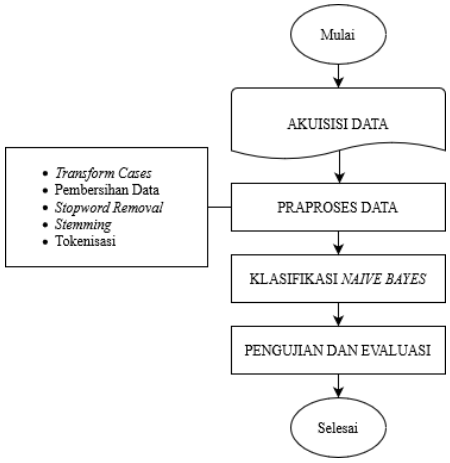
$$recall = \frac{TP}{TP+FN} \quad (8)$$

$$precision = \frac{TP}{TP+FP} \quad (9)$$

$$specificity = \frac{TN}{TN+FP} \quad (10)$$

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

3 METODOLOGI PENELITIAN



3.1 Akuisisi Data

Gambar. 1 Kerangka pikir di atas, berisi setiap proses yang telah dilakukan dalam mengerjakan penelitian ini.

Tahap ini merupakan tahap untuk mendapatkan data yang dibutuhkan, data yang dimaksud dalam penelitian ini adalah tweet pengguna twitter yang mention akun publik figur politik di Indonesia yang akan dipakai untuk data latih dan data uji. Data tweet didapatkan dengan cara menggunakan API(Application Programming Interface) yang sudah disediakan oleh twitter sebanyak 540 data tweet yang dibagi 440 data untuk data latih yang diberikan label positif dan negatif serta 100 tweet untuk data uji.

3.2 Praproses Data

Pada tahap ini sangat penting untuk mengurangi attribute pada data yang kurang berpengaruh pada proses klasifikasi dalam penelitian ini. Data yang digunakan pada langkah ini merupakan data data mentah yang memiliki noise, sehingga hasil dari tahap ini ialah data yang sudah siap untuk memudahkan proses klasifikasi. Pada praproses data terdapat beberapa tahap yaitu transform cases, pembersihan data, stopword removal, stemming, dan tokenisasi sehingga data yang masih terdapat noise akan bersih dan dapat digunakan untuk tahap selanjutnya

3.3 Klasifikasi Naïve Bayes

Tahap selanjutnya dari penelitian ini yaitu pengklasifikasian data tweet yang telah melewati tahap pra proses data metode yang digunakan yaitu klasifikasi dan penelitian ini menggunakan Naïve Bayes. Algoritma tersebut akan mengklasifikasikan tweet yang mengandung unsur cyberbullying. Dalam

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

pengerjaan menggunakan Naïve Bayes terdapat dua proses, yaitu data latih dan data uji. Langkah pertama ialah melakukan pelatihan sistem dengan data latih, Langkah selanjutnya yaitu proses uji sistem dengan mengacu probabilitas data latih. Klasifikasi Naïve Bayes menggunakan library dari NLTK (Natural Language Toolkit) merupakan *library* python yang digunakan dalam proses machine learning teks.

3.4 Pengujian Dan Evaluasi

Merupakan tahap untuk menguji hasil klasifikasi dengan menggunakan metode confusion matrix dengan sejumlah data yang diuji. Pada tahap ini menghitung nilai akurasi, precision, recall, dan spesificity.

4 Hasil Dan Pembahasan

Data yang digunakan pada penelitian ini sebanyak 540 data yang terbagi atas 440 data latih dan 100 data uji. Data latih terbagi atas 2 label yaitu data tweet positive yaitu data yang tidak mengandung unsur cyberbullying dan tweet negative merupakan data yang mengandung unsur cyberbullying, data latih tersebut diberikan label secara manual, dan data diambil dengan menggunakan kata kunci username twitter publik figur politik yang ada di Indonesia. Tabel II merupakan tabel yang berisi sampel dari data latih.

Tabel VII. Tabel Sampel Data Latih

Sampel Data Training	
Data Training	Label
alhamdulillah moga wabah virus covid cepat henti	Positif
keren gubernur Langkah cepat memvalidasi data terima bansos	Positif
tong kosong nyaring bunyi buzzer dasar monyet	Negatif
orang dungu asal posting buzzer	Negatif

Data tersebut kemudian dilakukan tahap praproses data, untuk membersihkan data dari noise sehingga data tersebut siap dilakukan proses klasifikasi, dan pada tahap akhir praproses data didapatkan hasil sebagai berikut.

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

Tabel VIII. Tabel Hasil Training Set

Term	Kelas		
	Positif	Negatif	
alhamdulillah	1	0	1/27
moga	1	0	1/27
wabah	1	0	1/27
virus	1	0	1/27
covid	1	0	1/27
cepat	2	0	2/27
henti	1	0	1/27
keren	1	0	1/27
gubernur	1	0	1/27
langkah	1	0	1/27
memvalidasi	1	0	1/27
data	1	0	1/27
terima	1	0	1/27
bansos	1	0	1/27
tong	0	1	1/27
kosong	0	1	

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

			1/27
nyaring	0	1	1/27
bunyi	0	1	1/27
buzzer	0	2	2/27
dasar	0	1	1/27
monyet	0	1	1/27
orang	0	1	1/27
dungu	0	1	1/27
asal	0	1	1/27
posting	0	1	1/27
	15/27	12/27	

Sampel data uji adalah “MasyaAllah.. Tenaga Medis yang meninggal dapat santunan dan anak"nya dapat beasiswa sampai kuliah!! Keren pak gubernur @aniesbaswedan” sehingga ketika dilakukan praproses data akan menghasilkan data uji seperti ‘masyaallah’ ‘tenaga’ ‘medis’ ‘tinggal’ ‘santun’ ‘anak’ ‘beasiswa’ ‘kuliah’ ‘keren’ ‘gubernur’.

Berdasarkan test set tadi, sistem harus membentuk instance dengan atribut yang sama dengan training set. Instance yang dibuat harus menggambarkan kedua hipotesis atau kelas yang tersedia.

Tahap selanjutnya yaitu menghitung probabilitas pada tiap kata yang ada di data uji yang mengacu data latih. Perhitungan dilakukan untuk setiap kemungkinan yang dapat terjadi, baik untuk hipotesis bernilai positif atau hipotesis yang bernilai negatif. Pada perhitungan ini, akan terdapat kata yang tidak pernah muncul sama sekali pada data latih. Untuk mencegah hal tersebut maka digunakan Laplace Smoothing.

Kemudian, tahap selanjutnya ialah untuk menghitung probabiltas class pada data uji tersebut, yang menghasilkan P(Positif|Datauji) sebesar 0,0000000000000001907 sedangkan P(Negatif|Datauji) sebesar 0,0000000000000001029, maka data uji tersebut diklasifikasikan masuk kedalam class positif.

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

Pada penelitian ini menggunakan data uji yang diambil pada tanggal 12 Mei 2020 pukul 01.00 dilakukan proses evaluasi dengan menggunakan confusion matrix untuk menghitung nilai akurasi, precision, recall, dan specificity mendapatkan hasil TP = 70, TN = 6, FP = 22, FN = 2

Tabel IV. Tabel Hasil Evaulasi dengan *Confusion Matrix*

Measure	Rumus	Perhitungan	Hasil
Akurasi	$\frac{TP+TN}{TP+FN+FP+TN}$	$\frac{70+6}{70+2+22+6}$	76%
Precision	$\frac{TP}{TP+FP}$	$\frac{70}{70+22}$	76,09%
Recall	$\frac{TP}{TP+FN}$	$\frac{70}{70+2}$	97,22%
spesificity	$\frac{TN}{TN+FP}$	$\frac{6}{6+22}$	21,4%

Nilai akurasi pada penelitian ini yaitu 76%, nilai precision sebesar 76,09%, nilai recall 97,22% sedangkan nilai specificity sebesar 21,4% Maka dengan dengan hasil evaluasi menggunakan confusion matrix ini dengan menggunakan data yang diambil pada tanggal 12 mei 2020 pukul 01.00 WIB untuk mengklasifikasi tweet yang mendung unsur cyber bullying ataupun tidak mengndung unsur cyber bullying masih kurang efektif, karena terdapat tweet yang terduplikat sehingga ketika program gagal melakukan proses klasifikasi pada tweet tersebut maka akan menggangu nilai evaluasi pada specificity, dan juga pada data latih negatif terdapat tweet negatif yang tidak mengandung unsur cyber bullying dan juga tweet negatif yang mengandung unsur cyberbullying.

5. Simpulan

Berdasarkan hasil penelitian yang diperoleh, dapat ditarik kesimpulan, yaitu algoritma Naïve bayes dapat digunakan untuk mengklasifikasikan tweet yang mengandung unsur cyberbullying maupun tidak mengandung unsur cyberbullying namun harus didukung dengan data latih yang bagus, baik untuk data latih dengan label yang mengandung unsur cyberbullying maupun data latih yang tidak mengandung unsur cyberbullying.
Pada evaluasi menggunakan confusion matrix didapatkan nilai akurasi sebesar 76%, precission 76,09%, recall 97,22% dan specificity sebesar 21,4%. Nilai specificity mendapatkan 21,4% dikarenakan data latih yang mengandung unsur cyerbullying masih tergolong kurang bagus, maka harus ditambahkan atau diperbaiki pada data latihnya, tetapi secara keseluruhan proses klasifikasi menggunakan algoritma naïve bayes ini berjalan dengan baik.

ISBN 978-623-93343-1-4
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 14 Agustus 2020

Referensi

[1] I. Y. U. R. Lingga, Pemodelan Deteksi Body Shaming Di Media Sosial Twitter Menggunakan Algoritma Naïve Bayes, vol. 8, no. 2. 2019.

[2] I. Y. Anggraini, S. Sucipto, and R. Indriati, "Cyberbullying Detection Modelling at Twitter Social Networking," *JUITA J. Inform.*, vol. 6, no. 2, p. 113, 2018, doi: 10.30595/juita.v6i2.3350.

[3] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," *Proc. 2nd Int. Conf. Knowl. Capture, K-CAP 2003*, no. January 2003, pp. 70–77, 2003, doi: 10.1145/945645.945658.

[4] L. F. S. Coletta, N. F. F. De Silva, E. R. Hruschka, and E. R. Hruschka, "Combining classification and clustering for tweet sentiment analysis," *Proc. - 2014 Brazilian Conf. Intell. Syst. BRACIS 2014*, no. January, pp. 210–215, 2014, doi: 10.1109/BRACIS.2014.46.

[5] D. Sarkar, *Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data*, vol. 32, no. 1. 2007.

[6] F. Sulianta, *Twitter For Business*. Elex Media Komputindo, 2013.

[7] R. B. Yates and B. R. Neto, "Modern Information Retrieval," no. January, 2015.

[8] C. D. Manning, J. Bauer, J. Finkel, and S. J. Bethard, "The Standford CoreNLP Natural Language Processing Toolkit," *Aclweb.Org*, pp. 55–60, 2014, [Online]. Available: http://macopolo.cn/mkpl/products.asp.

[9] D. L. Olson and D. Delen, *Advanced Data Mining Techniques. USA: Springer-Verlag Berlin Heidelberg*. Springer Science & Business Media, 2008.

[10] F. Handayani and S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *J. Tek. Elektro*, vol. 7, no. 1, pp. 19–24, 2015.

[11] M. C. Wijanto, "Sistem Pendeteksi Pengirim Tweet dengan Metode Klasifikasi Naive Bayes," *J. Tek. Inform. dan Sist. Inf.*, vol. 1, no. 2, pp. 172–182, 2015, doi: 10.28932/jutisi.v1i2.378.

[12] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeeding Eng.*, vol. 2, no. 1, pp. 1–7, 2015.