

## Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp

IRWANSYAH SAPUTRA  
DIDI ROSIYADI

Ilmu Komputer STMIK Nusa Mandiri  
Jalan Kramat Raya Nomor 18 Jakarta Pusat  
Email: 14002085@nusamandiri.ac.id, didi016@lipi.go.id

**Abstract.** WhatsApp is the most popular messaging application in Indonesia. This causes the emergence of cyberbullying behavior by its users. This study aims to classify WhatsApp chat to two classes, namely bully and not bully. The classification algorithms used are k-NN, NBC and SVM. The results show that the SVM algorithm is better at solving this case with an accuracy of 81.58%.

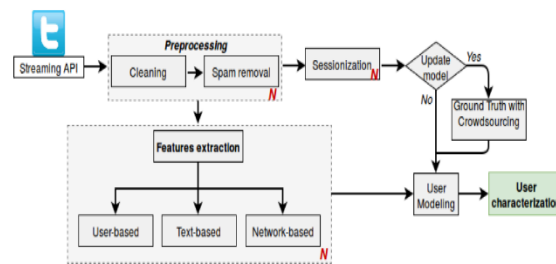
Key words: *cyberbullying, WhatsApp, k-NN, NBC, SVM*

**Abstrak.** WhatsApp merupakan aplikasi olah pesan terpopuler di Indonesia. Hal ini menyebabkan munculnya perilaku cyberbullying yang dilakukan para penggunanya. Penelitian ini bertujuan untuk mengklasifikasi teks obrolan WhatsApp kepada dua kelas, yaitu bully dan tidak bully. Algoritma klasifikasi yang digunakan yaitu k-NN, NBC dan SVM. Hasil menunjukkan bahwa algoritma SVM lebih baik dalam menyelesaikan kasus ini dengan akurasi sebesar 81,58%.

**Kata Kunci:** *cyberbullying, WhatsApp, k-NN, NBC, SVM*

### PENDAHULUAN

Pengguna internet di Indonesia lebih dari 132 juta jiwa pada tahun 2017 (Prabowo, 2017). Aplikasi olah pesan pada *smartphone* seperti WhatsApp *Messenger* pun mengalami peningkatan pengguna yang signifikan yaitu 35,8 juta pengguna dan menjadi aplikasi olah pesan terpopuler di Indonesia. Peningkatan penggunaan aplikasi ini menciptakan bentuk baru dari serangan yang disebut dengan *cyberbullying*. *Cyberbullying* merupakan suatu perilaku agresi yang mengacu pada perilaku *bullying* yang dilakukan oleh seseorang melalui sosial media seperti web, sms, jejaring sosial, *chat room*, dan lain-lain (Marcum, Higgins, Freiburger, & Ricketts, 2012). *Cyberbullying* menjadi masalah berbahaya karena memiliki dampak yang sangat serius terhadap psikis korban seperti perasaan sakit hati dan kecewa. *Cyberbullying* pada WhatsApp rentan dilakukan karena semakin banyaknya pengguna aplikasi tersebut yang berasal dari berbagai usia (Pingit, 2018). Riset sebelumnya melakukan deteksi agresi dan *bully* pada twitter berdasarkan fitur konten dan penyematan jaringan profil. Riset ini berfokus pada fenomena yang terjadi di media jejaring sosial Twitter. Media jejaring sosial ini memiliki beberapa hambatan untuk mendeteksi perilaku negatif yaitu karena *tweet* yang singkat, banyaknya *spam* dan tata bahasa rumit yang membuat lebih sulit pemrosesan mesin untuk memproses bahasa alami, mengekstrak atribut berbasis teks dan mengkarakterisasi interaksi pengguna Twitter. Pada penelitian ini, peneliti mengeksplorasi karakteristik pengguna Twitter sehubungan dengan konten dan penyematan jaringan seperti *following & follower* dan memanfaatkan atribut dengan *machine learning* klasifikasi untuk mendeteksi secara otomatis pengguna Twitter aggressor dan pengganggu (Chatzakou et al., 2017).



Gambar 1. Metode Pemecahan Masalah

Metode yang digunakan dalam analisis data, pelabelan dan klasifikasi hingga jutaan *tweet*, dan algoritma pemodelan yang dibangun *Random Forest classifier* dapat membedakan antara pengguna normal, agresif dan bullying dengan akurasi tinggi hingga >91% (Chatzakou et al., 2017).

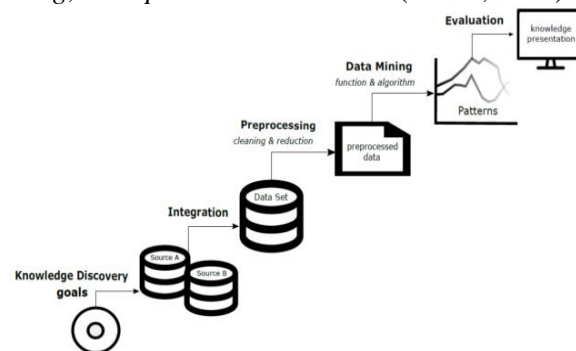
Selanjutnya terdapat penelitian yang dilakukan tentang pengenalan pernyataan sarkas pada WhatsApp Grup berbahasa Indonesia (Afiyati, Winarko, & Cherid, 2018). Sarkasme merupakan ekspresi kekesalan, kritik, ejekan dengan menggunakan kata-kata kasar yang dimaksudkan untuk menyinggung perasaan seseorang atau sesuatu. Contoh pernyataan sarkasme: "Saya tidak pernah melupakan wajah seseorang, tetapi dalam kasus Anda, saya akan senang membuat pengecualian" (Afiyati et al., 2018). Pernyataan sarkastik dapat menimbulkan kesulitan bagi banyak sistem berbasis *Natural Language Processing* (NLP). Teks data percakapan diambil dari grup WhatsApp menggunakan ponsel cerdas, dari menu Pengaturan - Obrolan Email - (pilih Tanpa Media) karena penelitian ini tidak akan membahas apa pun yang terkait dengan media, lalu pilih alamat email yang diinginkan untuk mengirim data. Data yang disediakan oleh WhatsApp adalah file teks. Data yang diambil memiliki jarak waktu tertentu, bergantung pada jumlah baris percakapan dan jumlah grup anggota sebagai kumpulan sampel. Selanjutnya dibangun vektor fitur untuk masing-masing contoh pernyataan dalam kumpulan sampel dan menggunakannya untuk membangun model klasifikasi. Fitur-fitur diekstraksi menggunakan berbagai komponen kata-kata, dan menghasilkan beberapa jenis sarkasme.

Kumpulan kata dan pernyataan yang dijalankan untuk eksperimen diperiksa dan diberi label secara manual. Metode yang diusulkan adalah metode berbasis pola. Metode ini menggunakan beberapa *set* fitur untuk mengklasifikasikan kalimat (pernyataan) kedalam dua bagian, kata positif dan kata negatif. Fitur kedua menggunakan pola tanda baca, jumlah tanda seru, tanda tanya, titik-titik, kata-kata kapital semua, tanda kutip dan jumlah vokal diulang lebih dari dua kali. Tanda tanya dan tanda seru dianggap sebagai indikator sarkasme yang mungkin. Tetapi tanda baca berfungsi sebagai prediktor terlemah karena jumlah tanda hanya ditemukan dalam jumlah yang sangat kecil. Fitur ketiga adalah fitur sintatik dan semantik. Fitur ini menghitung nilai dari kata-kata yang tidak umum (termasuk bahasa Inggris atau kata-kata bahasa lain kecuali bahasa Indonesia), kata-kata tidak umum seperti kata "heey" dalam pernyataan nomor 3, keberadaan ekspresi sarkastik umum, jumlah *interjections* dan ekspresi tertawa. Fitur-fitur ini menghasilkan prediktor yang lemah juga, karena tidak banyak tanda yang ditemukan. Hasil akurasi yang didapatkan tidak memadai karena penelitian ini hanya menggunakan data *sampling* dan eksperimen analisis secara manual (Afiyati et al., 2018).

## METODE

Metode penelitian yang digunakan pada penelitian ini menggunakan model KDD (*Knowledge Discovery in Database*). KDD adalah metode yang diusulkan oleh Fayyad pada tahun 1996. KDD merupakan proses mengekstraksi informasi baru dan pengetahuan dari *database* berukuran besar (Process, Williams, & Huang, 1987). Proses KDD menggunakan metode *data mining* untuk mengekstraksi apa yang dianggap pengetahuan sesuai dengan spesifikasi pengukuran dan ambang batas, penggunaan *database* dibutuhkan untuk melakukan *pre-processing* data, sub sampling dan transformasi dari *database* tersebut. Habibullah

menyatakan bahwa secara garis besar, metode KDD yang diusulkan oleh Fayyad memiliki 5 proses tahapan untuk sampai pada tujuannya yaitu *Knowledge Discovery Goals*, *Integration*, *Pre-processing Data mining*, *Interpretation/Evaluation* (Akbar, 2017).



Gambar 2. Alur Proses Metode KDD

### 1. *Knowledge Discovery Goals*

Tahapan pertama dalam mengerjakan sebuah aktifitas *data mining* adalah mengetahui tujuan organisasi berdasarkan permasalahan yang dimiliki. Penelitian ini bertujuan untuk mengklasifikasikan data uji ke dalam kelas yang sudah ditentukan, yaitu *bully* dan tidak *bully*. Proses pengujian dilakukan menggunakan tiga algoritma klasifikasi yaitu k-NN, NBC dan SVM.

### 2. *Data Integration*

Sesuai dengan tujuan kegiatan *data mining*, maka asal muasal sumber data akan ditentukan, dikumpulkan dan digabungkan menjadi sebuah data target. Jika domain aplikasi cukup besar, maka data target bisa berbentuk data *warehouse* ataupun data *mart*. Biasanya, tidak semua atribut data akan digunakan sehingga data dapat diseleksi dulu berdasarkan *subset* atribut yang relevan saja sehingga menjadi sebuah *dataset*.

*Dataset* yang digunakan pada penelitian ini memiliki jenis berbentuk teks obrolan pada grup WhatsApp “Pascasarjana 2017”. *Dataset* terdiri dari atribut teks dan kelas yang akan dilabeli menggunakan teknik *crowdsourcing* dengan 21 partisipan/responden. Label terdiri dari dua jenis, yaitu *bully* dan tidak *bully*.

### 3. *Data Preprocessing*

*Dataset* yang dihasilkan seringkali bersifat mentah dan kurang berkualitas, misal terdapat nilai yang hilang, salah input nilai, dan tidak konsisten. Akibatnya perlu dilakukan pembersihan data terlebih dahulu. Proses pembersihan mencakup menghilangkan duplikasi data, mengisi / membuang data yang hilang, memperbaiki data yang tidak konsisten, dan memperbaiki kesalahan ketik.

*Dataset* dibersihkan menggunakan teknik *Text Preprocessing* yang terdiri dari *tokenization* (menghilangkan simbol, tanda baca karakter khusus atau bukan huruf), *Normalization Indonesian Slang* (mengkonversi kata tidak baku menjadi baku), *stemming* (menghilangkan imbuhan), *stop word* (menghilangkan kata yang tidak bermakna, membuang data yang hilang, memperbaiki data yang tidak konsisten, dan memperbaiki kesalahan ketik) dan mentransformasi kata “tidak / bukan” untuk menghilangkan kata yang ambigu seperti kata tidak jelek, tidak serakah yang mengandung arti positif.

### 4. *Data mining*

Proses ini merupakan inti dari proses ekstraksi pengetahuan dari data. Algoritma / metode akan dipilih sesuai dengan tujuan kegiatan *data mining* yang telah ditentukan pada tahapan yang pertama.

Pengujian *dataset* pada penelitian ini dilakukan menggunakan tiga algoritma klasifikasi yaitu k-NN, NBC dan SVM yang akan diimplementasikan menggunakan tool RapidMiner versi 9.1.0.

### 5. *Knowledge Evaluation*

Penelitian ini membandingkan nilai tingkat akurasi antara pendekatan metode algoritma k-NN, NBC dan SVM yang dievaluasi menggunakan *precision*, *recall* dan *F-measure*. Penjelasan hasil evaluasi tersebut dapat diuraikan sebagai berikut (Maragoudakis, Fakotakis, & Kokkinakis, 2016)

a. *Precision*

*Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Formula *precision* adalah sebagai berikut:

$$p = \frac{tp}{tp+fp} \dots\dots\dots (1)$$

b. *Recall*

*Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Formula *Recall* adalah sebagai berikut:

$$r = \frac{tp}{tp+fn} \dots\dots\dots (2)$$

c. *F-measure*

*F-measure* merupakan salah satu perhitungan evaluasi dalam informasi temu kembali yang mengkombinasikan *recall* dan *precision*.

$$f = 1 / \left( a \frac{1}{p} + (1 - a) \frac{1}{r} \right) \dots\dots\dots (3)$$

## HASIL DAN PEMBAHASAN

### 1. *Knowledge Discovery Goals*

Penelitian ini bertujuan untuk mengklasifikasikan data uji kedalam kelas yang sudah ditentukan, yaitu *bully* dan tidak *bully*. *Cyberbullying* merupakan suatu perilaku agresi yang mengacu pada perilaku *bullying* yang dilakukan oleh seseorang melalui sosial media seperti web, sms, jejaring sosial, *chat room*, dan lain-lain. *Cyberbullying* menjadi masalah berbahaya karena memiliki dampak yang sangat serius terhadap psikis korban seperti perasaan sakit hati dan kecewa. Bentuk aktivitas *cyberbullying* meliputi berbagai jenis kegiatan negatif seperti *flaming* (mengirimkan pesan amarah, kasar dan vulgar), *harassment* (berulang kali mengirimkan pesan yang ofensif), *cyberstalking* (berulang kali mengirimkan ancaman membahayakan atau pesan-pesan yang sangat mengintimidasi), *denigration* (posting pernyataan yang tidak benar ataupun kejam), *impersonation* (berpura-pura menjadi orang lain untuk membuat orang tersebut terlihat buruk atau berada dalam bahaya), *outing* dan *trickery* (memposting hal-hal yang mengandung informasi pribadi atau *sensitive* mengenai orang lain atau memforward pesan-pesan pribadi dan atau terlibat dalam trik-trik dengan tujuan mengumpulkan informasi yang memalukan dan menyebarkannya), *exclusion* (dengan sengaja mengeluarkan seseorang dari *online group*). Walaupun *cyberbullying* tidak melibatkan kontak personal antara pelaku dan korban, tindakan ini dapat merusak psikologis dan emosional korban.

Penjelasan tersebut mengungkapkan pentingnya mendeteksi *cyberbullying* pada aplikasi olah pesan seperti WhatsApp agar dapat menimalisir jumlah korban *cyberbullying*. Proses pendeteksian *cyberbullying* dilakukan menggunakan tiga algoritma klasifikasi yaitu k-NN, NBC dan SVM.

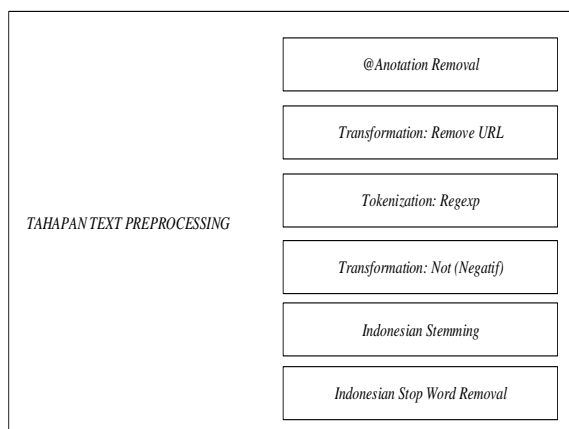
### 2. *Data Integration*

Tahapan ini menjelaskan asal sumber *dataset* dapat dikumpulkan. Data mentah yang akan dijadikan *dataset* memiliki jenis teks obrolan. Data tersebut dikumpulkan dari grup WhatsApp "Pascasarjana 2017". Terdapat 18 anggota aktif pada grup tersebut. Seluruh anggota grup merupakan mahasiswa pascasarjana STMIK Nusa Mandiri angkatan 2018-2019. *Dataset* terdiri dari atribut teks dan kelas dengan jumlah 1000 *record* yang akan dilabeli menggunakan teknik *crowdsourcing* dengan 21 partisipan/responden. Label terdiri dari dua bagian, yaitu *Bully* dan Tidak *bully*. Setelah mengalami pembersihan data, tersisa 380 *record* yang terdiri dari 204 teks

dengan kategori *bully* dan 176 *record* dengan kategori tidak. Seluruh *record* yang sudah diberi label akan dikelompokkan ke dalam satu *dataset* dalam bentuk comma separated value (\*.csv).

### 3. Data Preprocessing

Tahap ini bertujuan untuk membersihkan data dari *noise*. *Dataset* dibersihkan menggunakan teknik *Text Preprocessing* yang dilakukan menggunakan Gata *Framework* (Windu Gata, 2018) dan menggunakan fitur *Text Preprocessing* yang disediakan oleh RapidMiner. Tahapan *Text Preprocessing* yang digunakan dapat dilihat pada gambar 3.



Gambar 3. Tahapan *Text Preprocessing*

### 4. Tokenization

Tahapan tokenisasi merupakan proses yang dilakukan untuk menghilangkan simbol, tanda baca karakter khusus, karakter bukan huruf yang terdapat pada *dataset*.

Tabel 1. Perbandingan Saat Proses *Tokenization*

Proses	Teks
Teks sebelum dilakukan tokenisasi	Ya bagaimana baiknya, kalau ada masukan ya silahkan. Kalau mau rubah jadwal ya monggoh. Saya minggu tidak bisa kalau harus kuliah, karena minggu pekerjaan rumah sudah menunggu.
Teks setelah dilakukan tokenisasi	Ya bagaimana baiknya kalau ada masukan ya silahkan Kalau mau rubah jadwal ya monggoh Saya minggu tidak bisa kalau harus kuliah karena minggu pekerjaan rumah sudah menunggu

### 5. Normalization Indonesia Slang

Tahapan ini berfungsi untuk mengkonversi kata tidak baku menjadi baku. Seperti kata ngeselin arti sebenarnya adalah membuat kesal.

Tabel 2. Perbandingan Saat Proses *Normalization Indonesia Slang*

Proses	Teks
Teks sebelum dilakukan <i>Normalization Indonesia Slang</i>	Utk jurnal bisa dimana saja kampusnya Klo mau cari yg ga terlalu mahal coba di kampus swasta banyak tuh Utk periode february terbit
Teks setelah dilakukan <i>Normalization Indonesia Slang</i>	untuk membuat jurnal dapat memilih kampus apa saja Jika ingin mencari yang tidak terlalu mahal dapat mencoba di kampus-kampus swasta disana lebih banyak Untuk diterbitkan pada periode bulan february

### 6. Transformation Not (negatif)

Tahapan ini bertujuan untuk menghilangkan kata yang ambigu seperti kata tidak jelek, tidak serakah yang mengandung arti positif.

Tabel 3. Perbandingan Saat Proses *Transformation Not* (negatif)

Proses	Teks
Teks sebelum dilakukan <i>Transformation Not</i> (Negatif)	ya sudah kalau tidak sempat revisi Cek kembali format penulisannya Kirim kembali kalau sudah pasti Ditunggu
Teks setelah dilakukan <i>Transformation Not</i> (Negatif)	ya sudah kalau tidak sempat revisi Cek kembali format penulisannya Kirim kembali kalau sudah pasti Ditunggu

### 7. Stopword Removal

*Stopword Removal* merupakan tahapan untuk menghilangkan kata yang tidak relevan atau yang tidak bermakna. Kamus data menjadi sebuah acuan sebuah kata termasuk ke dalam *stop word* atau tidak. Kamus data tersebut berada pada program berbasis web yang dapat diakses *online* pada situs GataFramework (Windu Gata, 2018).

Tabel 4. Perbandingan Saat Proses *Stopword*

Proses	Teks
Teks sebelum dilakukan <i>Stopword</i>	mungkin pemerintah harus datang ke lapangan untuk mengetahui seperti apa kehidupan di lapangan jangan_diam saja di kursi empuk hanya bisa mengatur sana dan sini dan membuat peraturan yang tidak_sesuai dengan kenyataan
Teks setelah dilakukan <i>Stopword</i>	pemerintah lapangan kehidupan lapangan Jangan diam kursi empuk mengatur peraturan tidak_sesuai kenyataan

### 8. Stemming

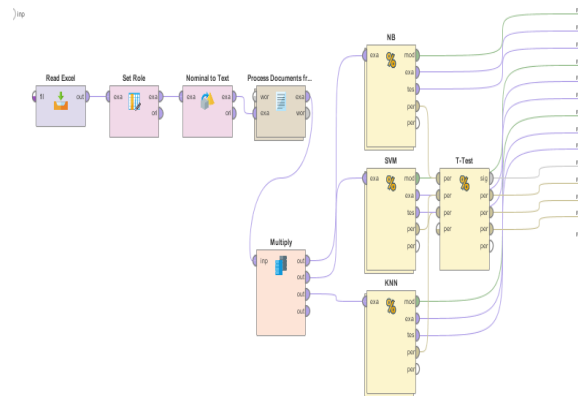
*Stemming* merupakan tahapan untuk mengembalikan suatu kata pada kata dasarnya (*root*) dengan istilah lain menghilangkan imbuhan yang terdapat pada kata tersebut.

Tabel 5. Perbandingan Teks saat dilakukan proses *Stemming*

Proses	Teks
Teks sebelum dilakukan proses <i>stemming</i>	semangat Masih seminggu dinilai diperbaiki kurangnya Kita buat buku Semuanya semangat lulus
Teks setelah dilakukan proses <i>stemming</i>	semangat masih minggu nilai baik kurang kita buat buku semua semangat lulus

### 9. Data mining

Tahapan ini melibatkan algoritma klasifikasi yang dibangun sesuai dengan tujuan awal penelitian. Pengujian *dataset* untuk mengklasifikasi kelas *bully* dan tidak *bully* pada penelitian ini dilakukan menggunakan tiga algoritma klasifikasi yaitu k-NN, NBC dan SVM yang diimplementasikan dengan tool RapidMiner versi 9.1.0. Desain RapidMiner yang diajukan adalah sebagai berikut,



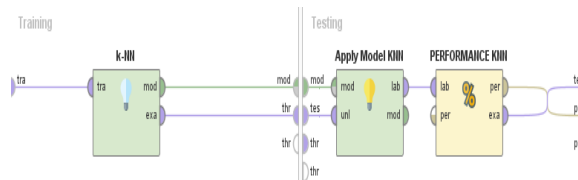
Gambar 4. Desain Model k-NN, NBC dan SVM

### 10. Knowledge Evaluation

Setelah model *data mining* didapatkan, proses selanjutnya adalah membandingkan nilai tingkat akurasi antara pendekatan metode algoritma k-NN, NBC dan SVM yang dievaluasi menggunakan *10-fold cross validation*.

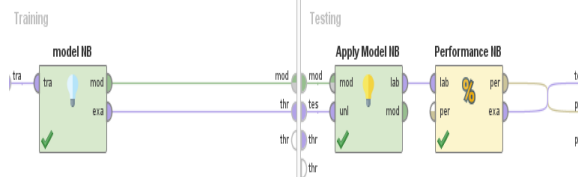
#### Hasil Pengujian Model Metode k-NN, NBC dan SVM

Desain proses pengujian model metode k-NN yang digunakan dapat dilihat pada gambar 5.



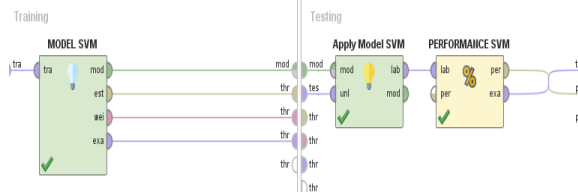
Gambar 5. Desain Model k-NN

Desain proses pengujian model NBC yang digunakan dapat dilihat pada gambar 6.



Gambar 6. Desain Model NBC

Desain proses pengujian model SVM yang digunakan dapat dilihat pada gambar 7.



Gambar 7. Desain Model SVM

Gambar 5, 6 dan 7 merupakan penjelasan detail dari desain proses yang terdapat pada gambar 4. Data yang digunakan untuk proses validasi terdiri dari dua jenis yaitu data dengan

kelas *bully* sebanyak 204 *record* dan data dengan kelas tidak *bully* sebanyak 176 *record*. Sebelum digunakan, seluruh data sudah melalui proses *Preprocessing* sehingga data tersebut bersih dari *noise* dan layak untuk digunakan. Jumlah *k* yang digunakan pada algoritma *k*-NN adalah *k*=3.

Selanjutnya, terdapat tabel *Confusion matrix* yang berisi informasi tentang klasifikasi aktual dan prediksi yang dilakukan oleh sistem klasifikasi. Kinerja sistem tersebut umumnya dievaluasi menggunakan data dalam matriks. Tabel berikut menunjukkan *Confusion matrix* pada klasifikasi yang memiliki kelas positif dan kelas negatif (Provost & Fawcett, 1997).

Tabel 6. *Confusion matrix*

		Actual	
		Negative	Positive
Predicted	Negative	<b>a</b>	<b>b</b>
	Positive	<b>c</b>	<b>d</b>

Keterangan:

a adalah jumlah prediksi yang benar bahwa sebuah instance negatif,

b adalah jumlah prediksi yang salah bahwa sebuah instance positif,

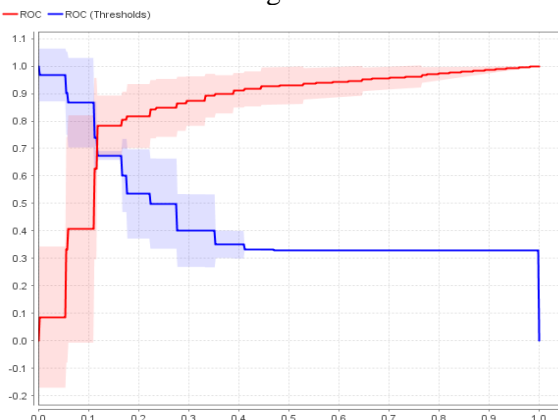
c adalah jumlah salah prediksi yang sebuah contoh negatif, dan

d adalah jumlah prediksi yang benar bahwa sebuah instance bernilai positif.

### Evaluasi Algoritma *k*-NN, NBC dan SVM

#### a. Kurva ROC

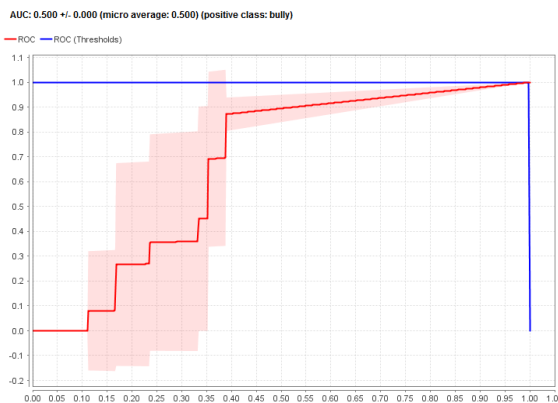
Penjelasan mengenai kurva ROC dari algoritma *k*-NN diuraikan sebagai berikut,



Gambar 8. Kurva ROC Algoritma *k*-NN

Kurva ROC algoritma *k*-NN dengan nilai *AUC* (*Area Under Curve*) yang dihasilkan dari gambar 8 sebesar 0,845. Hal tersebut berarti hasilnya adalah *good classification*.

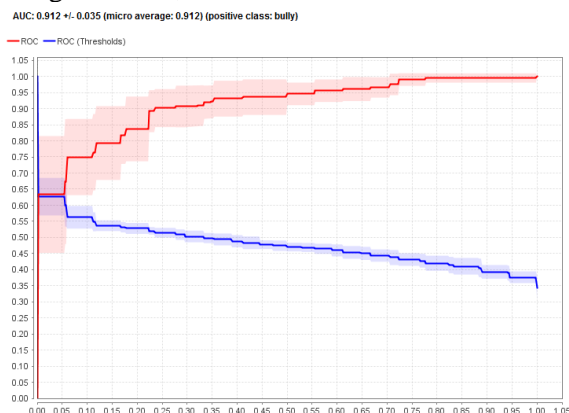
Selanjutnya, kurva ROC yang dimiliki algoritma NBC adalah sebagai berikut:



Gambar 9. Kurva ROC Algoritma NBC



Kurva ROC algoritma Naïve Bayes Classifier dengan nilai AUC (Area Under Curve) yang dihasilkan dari gambar 9 sebesar 0,500. Kemudian, kurva ROC yang dimiliki oleh algoritma SVM adalah sebagai berikut,



Gambar 10. Kurva ROC Algoritma SVM

Kurva ROC algoritma Support Vector Machine dengan nilai AUC (Area Under Curve) yang dihasilkan dari gambar 10 sebesar 0,912.

**b. Confusion matrix**

Selain kurva ROC, pengukuran akurasi juga dilakukan menggunakan Confusion matrix. Hasil perhitungan akurasi untuk algoritma k-NN dapat dilihat pada tabel 7.

Tabel 7. Confusion matrix algoritma k-NN

Accuracy: 81,32% +/- 6,91%			
	True Tidak	True Bully	Class Precision
Pred. Tidak	140	35	80,00%
Pred. Bully	36	169	82,44%
Class Recall	79,55%	82,84%	

Akurasi yang diperoleh adalah sebesar 81,32% dari 204 data kelas bully dan 176 data kelas tidak bully pada teks obrolan grup “Pascasarjana 2017” pada aplikasi WhatsApp Messenger.

Data kelas bully yang sesuai dengan prediksi yaitu sebanyak 169 record. Data tidak bully yang termasuk ke dalam prediksi Tidak yaitu 35 record. Data tidak bully yang termasuk ke dalam prediksi bully yaitu 36 record dan data tidak bully yang sesuai prediksi sebanyak 140 record.

Selanjutnya, hasil perhitungan akurasi untuk algoritma NBC dapat dilihat pada tabel 8.

Tabel 8. Confusion matrix algoritma NBC

Accuracy: 78,95% +/- 6,45%			
	True Tidak	True Bully	Class Precision
Pred. Tidak	126	30	80,77%
Pred. Bully	50	174	77,68%
Class Recall	71,59%	85,29%	

Akurasi yang diperoleh adalah sebesar 78,95% dari 204 data kelas bully dan 176 data tidak bully pada teks obrolan grup “Pascasarjana 2017” pada aplikasi WhatsApp Messenger.

Data kelas bully yang sesuai dengan prediksi yaitu sebanyak 174 record. Data tidak bully yang termasuk ke dalam prediksi Tidak yaitu 30 record. Data tidak bully yang termasuk ke

dalam prediksi *bully* yaitu 50 *record* dan data tidak *bully* yang sesuai prediksi sebanyak 126 *record*.

Terakhir, hasil perhitungan akurasi untuk algoritma SVM dapat dilihat pada tabel 9.

Tabel 9. *Confusion matrix* algoritma SVM

<i>Accuracy: 81,58% +/- 4,24%</i>			
	<i>True Tidak</i>	<i>True Bully</i>	<i>Class Precision</i>
<i>Pred. Tidak</i>	125	19	86,81%
<i>Pred. Bully</i>	51	185	78,39%
<i>Class Recall</i>	71,02%	90,69%	

Akurasi yang diperoleh adalah sebesar 81,58% dari 204 data kelas *bully* dan 176 data tidak *bully* pada teks obrolan grup “Pascasarjana 2017” pada aplikasi WhatsApp Messenger.

Data kelas *bully* yang sesuai dengan prediksi yaitu sebanyak 185 *record*. Data tidak *bully* yang termasuk ke dalam prediksi Tidak yaitu 19 *record*. Data tidak *bully* yang termasuk ke dalam prediksi *bully* yaitu 51 *record* dan data tidak *bully* yang sesuai prediksi sebanyak 125 *record*.

### c. *Precision & Recall*

*Precision & Recall* dari algoritma k-NN, NBC dan SVM dapat dilihat pada tabel 10.

Tabel 10. *Precision & Recall* algoritma k-NN, NBC & SVM

Algoritma	<i>Precision</i>	<i>Recall</i>
k-NN	82,46%	82,83%
NBC	78,08%	85,38%
SVM	78,55%	90,69%

### d. *F-measure*

*F-measure* dari algoritma k-NN, NBC dan SVM dapat dilihat pada tabel 11.

Tabel 11. *F-measure* algoritma k-NN, NBC & SVM

Algoritma	<i>F-measure</i>
k-NN	82,47%
NBC	81,38%
SVM	84,03%

Setelah pengujian selesai dilakukan, perbandingan akurasi dari kinerja algoritma k-NN, NBC dan SVM dapat dilihat pada tabel 12.

Tabel 12. Perbandingan Akurasi dan AUC Ketiga Algoritma

Algoritma	Akurasi	AUC
k-NN	81,32%	0,845
NBC	78,95%	0,500
SVM	<b>81,58%</b>	<b>0,912</b>

Tabel 12 menunjukkan bahwa akurasi yang dimiliki oleh algoritma SVM sebesar 81,58% lebih baik dibandingkan dengan kinerja algoritma k-NN sebesar 81,32% dan algoritma NBC sebesar 78,95%.

## PENUTUP

### Simpulan

Peningkatan penggunaan aplikasi WhatsApp menciptakan bentuk baru dari serangan yang disebut dengan cyberbullying. Penelitian ini menggunakan fitur yang disediakan WhatsApp untuk mengekspor percakapan grup. Dataset merupakan teks obrolan dari grup “Pascasarjana 2017”. Setelah dataset ditentukan, proses selanjutnya adalah melakukan pembersihan data noise atau derau menggunakan Text Preprocessing, kemudian dilakukan pelabelan pada kelas *bully* dan tidak *bully* menggunakan teknik crowdsourcing dengan 21 partisipan. Selanjutnya, melakukan proses klasifikasi terhadap dataset tersebut dengan cara

membandingkan kinerja antara tiga algoritma yaitu k-NN, NBC dan SVM dalam penyelesaian masalah klasifikasi tingkah laku bully pada teks obrolan di aplikasi WhatsApp Messenger. Hasil pengujian menunjukkan bahwa akurasi yang dimiliki oleh algoritma k-NN, NBC dan SVM yaitu masing-masing sebesar 81.32%, 78.95% dan 81.58%. Pengujian tersebut dapat diketahui bahwa algoritma SVM lebih baik dibanding kedua algoritma lainnya.

### Saran

Penggunaan atribut lainnya seperti atribut nama user dan waktu obrolan dikirim. “Nama User” untuk mendapatkan siapa yang paling sering membully pada grup tersebut. Sedangkan atribut “waktu obrolan dikirim” digunakan untuk mendapatkan intensitas pembullying dilakukan antar anggota grup tersebut. Atribut ini juga dapat dipakai untuk mengetahui jam atau hari intensitas terbanyak dilakukannya pembullying.

### DAFTAR PUSTAKA

- Afiyati, R., Winarko, E., & Cherid, A. (2018). Recognizing the sarcastic statement on WhatsApp Group with Indonesian language text. *2017 International Conference on Broadband Communication, Wireless Sensors and Powering, BCWSP 2017, 2018–Janua*(May), 1–6. <https://doi.org/10.1109/BCWSP.2017.8272579>
- Akbar, H. (2017). Ingin Terapkan Data Mining? Ini Tahapannya. Retrieved from <https://mti.binus.ac.id/2017/12/05/ingin-terapkan-data-mining-ini-tahapannya/>
- Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017). Detecting Aggressors and Bullies on Twitter. In *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion* (pp. 767–768). <https://doi.org/10.1145/3041021.3054211>
- Maragoudakis, M., Fakotakis, N., & Kokkinakis, G. (2016). A Bayesian Model for Shallow Syntactic Parsing of Natural Language Texts, (January 2016).
- Marcum, C. D., Higgins, G. E., Freiburger, T. L., & Ricketts, M. L. (2012). Battle of the sexes: An examination of male and female cyber bullying. *International Journal of Cyber Criminology*, 6(1), 904–911.
- Pingit, A. (2018). WhatsApp Naikkan Batas Usia Pengguna Menjadi 16 Tahun. Retrieved from <https://katadata.co.id/berita/2018/04/27/whatsapp-naikkan-batas-usia-pengguna-dari-menjadi-16-tahun>
- Prabowo, A. (2017). Pengguna Ponsel Indonesia Mencapai 142% dari Populasi. Retrieved from <https://databoks.katadata.co.id/datapublish/2017/08/29/pengguna-ponsel-indonesia-mencapai-142-dari-populasi>
- Process, F. S., Williams, G. J., & Huang, Z. (1987). Modelling the KDD Process. *Proc Centre for Software Reliability Conference on Measurement for Software Control and Assurance*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.43.4&rep=rep1&type=pdf>
- Provost, F., & Fawcett, T. (1997). Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions. In and R. U. David Heckerman, Heikki Mannila, Daryl Pregibon (Ed.), *THE THIRD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING* (pp. 43–48). Newport Beach, California, USA: AAAI.
- Saputra, I., & Rosiyadi, D. (2019). Laporan Akhir Penelitian.
- Windu Gata. (2018). Text Mining Program. Retrieved from <http://www.gataframework.com/textmining>