

**Perbandingan Kinerja Algoritma *k-Nearest Neighbor*, *Naïve Bayes Classifier*
dan *Support Vector Machine* dengan *SMOTE* untuk Klasifikasi Tingkah
Laku *Bully* pada Aplikasi *Messenger WhatsApp***



TESIS

Irwansyah Saputra
14002085

PROGRAM STUDI MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2019

**Perbandingan Kinerja Algoritma *k-Nearest Neighbor*, *Naïve Bayes Classifier*
dan *Support Vector Machine* dengan *SMOTE* untuk Klasifikasi Tingkah
Laku *Bully* pada Aplikasi *Messenger WhatsApp***



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Magister Ilmu Komputer (M.Kom)

Irwansyah Saputra
14002085

PROGRAM STUDI MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2019

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama : Irwansyah Saputra
NIM : 14002085
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Data Mining*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: “Perbandingan Kinerja Algoritma *k-Nearest Neighbor*, *Naïve Bayes Classifier* dan *Support Vector Machine* dengan *SMOTE* untuk Klasifikasi Tingkah Laku *Bully* pada Aplikasi *Messenger WhatsApp*” adalah hasil karya sendiri dan semua sumber baik yang kutip maupun yang rujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 8 Januari 2019
Yang menyatakan,

Materai Rp. 6.000,-

Irwansyah Saputra

HALAMAN PENGESAHAN

Tesis ini diajukan oleh:

Nama : Irwansyah Saputra
NIM : 14002085
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Data Mining*
Judul Tesis : “Perbandingan Kinerja Algoritma *k-Nearest Neighbor*, *Naïve Bayes Classifier* dan *Support Vector Machine* dengan *SMOTE* untuk Klasifikasi Tingkah Laku *Bully* pada Aplikasi *Messenger WhatsApp*”

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Studi Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 26 Januari 2019
Program Studi Magister Ilmu Komputer
STMIK Nusa Mandiri
Ketua

Dr. Dwiza Riyana, S.Si, MM, M.Kom.

DEWAN PENGUJI

Penguji I :

Penguji II :

Penguji III / Pembimbing : Dr. Didi Rosiyadi

KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut “Perbandingan Kinerja Algoritma *k-Nearest Neighbor*, *Naïve Bayes Classifier* dan *Support Vector Machine* dengan *SMOTE* untuk Klasifikasi Tingkah Laku *Bully* pada Aplikasi *Messenger WhatsApp*”.

Tujuan penulisan tesis ini dibuat sebagai salah satu untuk mendapatkan gelar Magister Ilmu Komputer (M.Kom) pada Program Studi Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (PPs MIK STMIK Nusa Mandiri).

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu izinkanlah penulis pada kesempatan ini untuk mengucapkan ucapan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Didi Rosiyadi, M.Kom. selaku pembimbing tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis dalam menyelesaikan tesis ini.
2. Orang tua tercinta, Mulya Warni dan Ujang Maman yang telah memberikan dukungan material dan moral kepada penulis.
3. Adik-adik tercinta, Muhammad Nauval Azhar, Laila Tazkiyah dan Amanda Aprilia Azmi.
4. Seorang terbaik yang selalu hadir dan menyemangati dari Strata 1 hingga sekarang, T.Z.F.
5. Seluruh staf pengajar (dosen) PPs MIK STMIK Nusa Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
6. Seluruh staf dan karyawan PPs MIK STMIK Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.
7. Teman-teman seperjuangan di tempat kuliah.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 26 Januari 2019

Irwansyah Saputra

Penulis

**SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini, saya:

Nama : Irwansyah Saputra
NIM : 14002085
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Data Mining*
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan izin kepada pihak Program Studi Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : “Perbandingan Kinerja Algoritma *k-Nearest Neighbor*, *Naïve Bayes Classifier* dan *Support Vector Machine* dengan *SMOTE* untuk Klasifikasi Tingkah Laku *Bully* pada Aplikasi *Messenger WhatsApp*” beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta izin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 26 Januari 2019
Yang menyatakan,

Materai Rp. 6.000,-

Irwansyah Saputra

ABSTRAK

Nama : Irwansyah Saputra
NIM : 14002085
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Data Mining*
Judul : “Perbandingan Kinerja Algoritma *k-Nearest Neighbor*, *Naïve Bayes Classifier* dan *Support Vector Machine* dengan *SMOTE* untuk Klasifikasi Tingkah Laku *Bully* pada Aplikasi *Messenger WhatsApp*”

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. Hasil pengujian tanpa *SMOTE* 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. Pengujian selanjutnya yang ditambahkan *SMOTE* untuk algoritma k-NN, NBC dan SVM yaitu sebesar 78.91%, 82.59% dan 83.57% menunjukkan bahwa algoritma SVM merupakan algoritma terbaik dengan *SMOTE* maupun tidak dalam menyelesaikan masalah klasifikasi tingkah laku *bully* pada teks obrolan aplikasi WhatsApp *Messenger*.

Kata kunci:

cyberbullying, WhatsApp Messenger, k-NN, SVM, NBC, SMOTE

ABSTRACT

Name : Irwansyah Saputra
NIM : 14002085
Study of Program : Master of Computer Science
Levels : Strata Dua (S2)
Concentration : *Data Mining*
Title : “Perbandingan Kinerja Algoritma *k-Nearest Neighbor*, *Naïve Bayes Classifier* dan *Support Vector Machine* dengan *SMOTE* untuk Klasifikasi Tingkah Laku *Bully* pada Aplikasi *Messenger WhatsApp*”

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 text to two classes, namely bully and not bully. The test results without SMOTE showed that the accuracy possessed by the k-NN, NBC and SVM algorithms was 81.32%, 78.95% and 81.58% respectively. The test can be seen that the SVM algorithm is better than the other two algorithms. The next test added by SMOTE for the k-NN, NBC and SVM algorithms which is 78.91%, 82.59% and 83.57% indicates that the SVM algorithm is the best algorithm with SMOTE or not in solving bully behavior classification problems in the text of WhatsApp Messenger application chat.

Keywords:
cyberbullying, WhatsApp Messenger, k-NN, SVM, NBC, SMOTE

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
SURAT PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR	v
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	vii
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I	1
PENDAHULUAN	1
2.1. Latar Belakang Penulisan	1
2.2. Masalah Penelitian	5
1.2.1. Identifikasi Masalah	5
1.2.2. Batasan Masalah.....	5
1.2.3. Rumusan Masalah	6
2.3. Tujuan Penelitian.....	6
1.4. Ruang Lingkup Penelitian	7
1.5. Hipotesis	7
1.6. Sistematika Penulisan.....	7
BAB II.....	9
LANDASAN KERANGKA PEMIKIRAN	9
2.4. Tinjauan Pustaka	9
2.1.1. Data mining.....	9
2.1.2. Fungsi Data mining.....	10
2.1.3. Klasifikasi	12
2.1.4. Text Mining.....	13
2.1.4.1. Tahapan Proses Text Mining (Text preprocessing).....	13
2.1.4.2. Gata Framework	16
2.1.5. Support Vector Machine (SVM).....	16
2.1.6. Naive Bayes Classifier (NBC)	18
2.1.7. k-Nearest Neighbor (k-NN)	21
2.1.8. Seleksi Fitur SMOTE	24
2.1.9. Crowdsourcing	26

2.1.10.	Cyberbullying.....	28
2.2.	Tinjauan Studi	29
2.3.	Kerangka Pemikiran	43
2.4.	Objek Penelitian	44
BAB III		45
METODOLOGI PENELITIAN.....		45
3.1.	Metode Penelitian.....	45
3.1.1.	Knowledge Discovery Goals.....	46
3.1.2.	Data Integration.....	46
3.1.3.	Data Preprocessing.....	47
3.1.4.	Data mining.....	49
3.1.5.	Knowledge Evaluation	51
BAB IV		52
HASIL PENELITIAN DAN PEMBAHASAN		52
4.1.	KDD (<i>Knowledge Discovery in Database</i>).....	52
4.1.1.	Knowledge Discovery Goals.....	52
4.1.2.	Data Integration.....	53
4.1.3.	Data Preprocessing.....	54
4.1.3.1.	Tokenization	55
4.1.3.2.	Normalization Indonesia Slang.....	58
4.1.3.3.	Transformation Not (negatif).....	60
4.1.3.4.	Stopword Removal	62
4.1.3.5.	Stemming.....	65
4.1.4.	Data mining.....	67
4.1.5.	Knowledge Evaluation	68
4.1.5.1.	Hasil Pengujian Model Metode k-NN	68
4.1.5.2.	Hasil Pengujian Model Metode Naïve Bayes Classifier.....	69
4.1.5.3.	Hasil Pengujian Model Metode Support Vector Machine	69
4.1.5.4.	Evaluasi Algoritma k-NN	70
4.1.5.5.	Evaluasi Algoritma Naïve Bayes Classifier.....	72
4.1.5.6.	Evaluasi Algoritma Support Vector Machine.....	74
4.1.5.7.	Evaluasi Algoritma k-NN + SMOTE	77
4.1.5.8.	Evaluasi Algoritma Naïve Bayes Classifier + SMOTE.....	78
4.1.5.9.	Evaluasi Algoritma Support Vector Machine + SMOTE.....	80
4.1.5.10.	Evaluasi Algoritma k-NN, NBC dan SVM	83
BAB V.....		88
PENUTUP.....		88
5.1.	Kesimpulan.....	88
5.2.	Saran.....	88
DAFTAR PUSTAKA		93

DAFTAR TABEL

Tabel 2.1. Performa kombinasi algoritma klasifikasi (Li & Jain, 1998).....	31
Tabel 2.2. Hasil Pengujian algoritma SVM (Sarkar et al., 2015)	36
Tabel 2.3 Contoh Pernyataan	38
Tabel 2.4 Ringkasan Penulisan	39
Tabel 4.1. Daftar Nama Anggota Grup WhatsApp Pascasarjana 2017	54
Tabel 4.2. Perbandingan Teks Saat Dilakukan Proses Tokenization.....	55
Tabel 4.3. Perbandingan Teks saat dilakukan proses Normalization Indonesia Slang.....	58
Tabel 4.4. Perbandingan Teks saat dilakukan proses Transformation Not (negatif)	60
Tabel 4.5. Perbandingan Teks saat dilakukan proses Stop Word	62
Tabel 4.6. Perbandingan Teks saat dilakukan proses Stemming	65
Tabel 4.7. Confusion matrix (Provost & Fawcett, 1997).....	70
Tabel 4.8. Confusion matrix Algoritma k-NN	71
Tabel 4.9. Precision & Recall Algoritma k-NN.....	72
Tabel 4.10. F-Measure Algoritma k-NN.....	72
Tabel 4.11. Confusion matrix Algoritma Naïve Bayes Classifier	73
Tabel 4.12. Precision & Recall Algoritma Naïve Bayes Classifier	74
Tabel 4.13. F-Measure Algoritma Naïve Bayes Classifier	74
Tabel 4.14. Confusion matrix Algoritma Naïve Bayes Classifier	75
Tabel 4.15. Precision & Recall Algoritma Support Vector Machine	76
Tabel 4.16. F-Measure Algoritma Support Vector Machine	76
Tabel 4.17. Perbandingan Akurasi dan AUC Ketiga Algoritma.....	76
Tabel 4.18. Confusion matrix Algoritma k-NN + SMOTE	78
Tabel 4.19. Precision & Recall Algoritma k-NN.....	78
Tabel 4.20. F-Measure Algoritma k-NN.....	78
Tabel 4.21. Confusion matrix Algoritma Naïve Bayes Classifier	79
Tabel 4.22. Precision & Recall Algoritma Naïve Bayes Classifier	80
Tabel 4.23. F-Measure Algoritma Naïve Bayes Classifier	80
Tabel 4.24. Confusion matrix Algoritma Support Vector Machine + SMOTE....	81
Tabel 4.25. Precision & Recall Algoritma Support Vector Machine	82
Tabel 4.26. F-Measure Algoritma Support Vector Machine	82
Tabel 4.27. Perbandingan Akurasi dan AUC dari Ketiga Algoritma	82
Tabel 4.28. Confusion matrix Algoritma k-NN, NBC dan SVM	83
Tabel 4.29. Precision & Recall Algoritma k-NN, NBC dan SVM	84
Tabel 4.30. F-Measure Algoritma k-NN, NBC dan SVM.....	84
Tabel 4.31. T-Test Algoritma k-NN, NBC dan SVM.....	85
Tabel 4.32. T-Test Algoritma k-NN, NBC dan SVM dengan fitur SMOTE.....	85

DAFTAR GAMBAR

Gambar 1.1. Top 10 Aplikasi Mobile di Indonesia (Hadi Pratama, 2017)	1
Gambar 2.1 Clustering (MacLennan et al., 2008).....	11
Gambar 2.2 Product Association (MacLennan et al., 2008)	11
Gambar 2.3 web Navigation Sequence (MacLennan et al., 2008).....	12
Gambar 2.4 Proses Case Folding	14
Gambar 2.5 Proses Tokenizing	15
Gambar 2.6 Proses Stemming	15
Gambar 2.7 SVM berusaha menemukan hyperplane terbaik yang memisahkan kedua class -1 dan +1 (Cortes & Vapnik, 1995)	18
Gambar 2.8 Contoh klasifikasi k-NN (Jadhav & Channe, 2013)	22
Gambar 2.9 Konsep Seleksi Fitur SMOTE.....	25
Gambar 2.10 Sampel Data yang Tidak Seimbang	25
Gambar 2.11. Data Minority Dihubungkan	26
Gambar 2.12. Menambahkan Data Acak	26
Gambar 2.13. Data Sintetik Baru yang Ditambahkan.....	26
Gambar 2.14 Hasil perbandingan empat algoritma klasifikasi	30
Gambar 2.15 Total Recall, Precision & F-Measure (Rasjid & Setiawan, 2017) ..	32
Gambar 2.16 Hasil Akurasi k-NN (Rasjid & Setiawan, 2017)	33
Gambar 2.19 Metode Pemecahan Masalah (Chatzakou et al., 2017)	34
Gambar 3.1 Alur Proses Metode KDD (Akbar, 2017).....	45
Gambar 3.2 Model yang Diusulkan	48
Gambar 3.3 Model Diagram Alur Aplikasi.....	50
Gambar 4.1 Tahapan Text Preprocessing	55
Gambar 4.2 Desain Model k-NN, NB dan SVM	67
Gambar 4.3. Desain Model k-NN, NB dan SVM dengan fitur SMOTE	68
Gambar 4.4. Desain Model Metode k-NN	68
Gambar 4.5. Desain Model Metode Naïve Bayes Classifier	69
Gambar 4.6. Desain Model Metode Support Vector Machine.....	69
Gambar 4.7 Gambar Kurva ROC Algoritma k-NN	71
Gambar 4.8. Gambar Kurva ROC Algoritma Naïve Bayes Classifier.....	73
Gambar 4.9. Gambar Kurva ROC Algoritma Support Vector Machine.....	75
Gambar 4.10 Gambar Kurva ROC Algoritma k-NN + SMOTE	77
Gambar 4.11. Gambar Kurva ROC Algoritma Naïve Bayes Classifier.....	79
Gambar 4.12. Gambar Kurva ROC Algoritma SVM + SMOTE.....	81
Gambar 4.13. Menu Login	86
Gambar 4.14. Menu Input Teks	86
Gambar 4.15. Form Text preprocessing dan Hasil SVM.....	87

DAFTAR LAMPIRAN

Lampiran 1. Lembar Bimbingan Tesis.....	98
Lampiran 2. Surat Izin Pengambilan Data	99
Lampiran 3. Surat Persetujuan Sebagai Responden Penelitian.....	100

PENDAHULUAN

2.1. Latar Belakang Penulisan

Pengguna internet di Indonesia lebih dari 132 juta jiwa pada tahun 2017 (Prabowo, 2017). Hal ini berbanding lurus dengan pengguna *smartphone* yang diperkirakan akan meningkat lebih dari 100% dibanding tahun sebelumnya (Rahmayani, 2015). 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 dibandingkan dengan jumlah pengguna aplikasi BBM sebanyak 34,7 juta dan aplikasi Line sebanyak 27,6 Juta pengguna (Hadi Pratama, 2017).



Gambar 1.1. Top 10 Aplikasi Mobile di Indonesia (Hadi Pratama, 2017)

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 (Satalina, 2014). UNICEF merilis laporan tahun 2015 bahwa terdapat 40% anak Indonesia mengalami peristiwa *bully* (UNICEF, 2015). Para peneliti bidang ilmu psikologi menyatakan bahwa *bully* termasuk atau memiliki kesamaan perbuatan dengan agresi (W Rahmawati, 2016). *Cyberbullying* pada WhatsApp rentan dilakukan karena semakin banyaknya pengguna aplikasi tersebut yang berasal dari berbagai usia (Pingit, 2018). WhatsApp memiliki fitur *end to end encryption* atau E2EE, yaitu setiap pesan yang dikirim langsung dienkripsi secara aman dan hanya dapat dibuka oleh pengirim dan penerima saja (WhatsApp Inc., 2018). Penelitian ini akan menggunakan fitur yang disediakan WhatsApp untuk mengekspor percakapan grup karena *smartphone* yang dipakai dalam keadaan normal. Setelah data diekspor, proses selanjutnya adalah melakukan pembersihan *data noise* atau derau menggunakan *Text preprocessing* seperti *@anotation removal* yaitu menghapus simbol anotasi, *transformation: Remove URL* untuk menghapus link dalam data, *tokenization: Regexp* yaitu tahap pemotongan *string* masukan berdasarkan kata yang menyusunnya, *transformation: not (negative)* untuk memastikan kata yang ambigu seperti kata tidak jelek, tidak serakah yang mengandung arti positif, *Normalization Indonesian Slang* yaitu mengkonversi kata tidak baku menjadi kata baku, selanjutnya proses *stemming* yaitu mencari kata dasar sekaligus menghilangkan imbuhan dari kata tersebut, terakhir adalah proses *stopword* yaitu tahap pengambilan kata-kata penting hasil dari *token* dan membuang kata-kata yang tidak bermanfaat (Vijayarani et al., 2015). Selanjutnya, data mentah yang sudah dibersihkan akan diberi label menggunakan teknik *crowdsourcing*, yaitu teknik yang memanfaatkan sumberdaya dari banyak orang untuk melakukan suatu pekerjaan dalam kasus ini kegiatan tersebut adalah menentukan *class dataset* (Howe, 2009). Jumlah partisipan tidak ditentukan dengan suatu aturan. Penelitian ini menggunakan 21 partisipan yang dipilih untuk menentukan *class bully* dan tidak *bully*. Setelah pembersihan data dan pelabelan selesai, tahapan selanjutnya adalah melakukan proses klasifikasi terhadap *dataset* tersebut dengan cara membandingkan kinerja antara tiga algoritma yaitu *Naïve Bayes Classifier* (NBC) dan *k-Nearest Neighbor* (k-NN) dan *Support Vector Machine* (SVM) yang diberi tambahan *SMOTE* dalam penyelesaian masalah

klasifikasi tingkah laku *bully* pada teks obrolan di aplikasi WhatsApp Messenger. Pengklasifikasian tersebut akan menghasilkan evaluasi akurasi, AUC, *precision*, *recall* dan *F-Measure*.

Algoritma NBC dan k-NN dan SVM merupakan beberapa algoritma yang populer untuk menangani masalah klasifikasi, bahkan ketiga algoritma ini masuk kedalam top 10 algoritma terbaik versi IEEE Konferensi Internasional tentang *Data mining* pada Desember 2006 (Wu et al., 2008). Pernyataan tersebut diperkuat dengan berbagai penelitian yang membahas tentang algoritma-algoritma tersebut dalam menangani klasifikasi, khususnya klasifikasi berbasis teks. Haiyi Zhang dan Di Li mengungkapkan bahwa algoritma *Naïve Bayes* merupakan algoritma yang sangat sederhana namun menunjukkan kelebihan dalam teks klasifikasi serta memberikan hasil yang memuaskan dalam menyelesaikan kasus klasifikasi *email spam* (H. Zhang & Li, 2007). Penelitian lainnya dilakukan Hassan, Rafi dan Shahid Shaikh yang membahas tentang pelabelan dokumen menggunakan pengetahuan repositori eksternal seperti *Word Net*, *Open Project Directory (OPD)*, *Wikipedia* dan *Wikitology*. Hasilnya menunjukkan bahwa *Wikitology* merupakan basis pengetahuan yang lebih baik dibanding yang lainnya. Hasil tersebut didapatkan dari proses klasifikasi menggunakan algoritma SVM dan NBC dengan 10 kali lipat *cross validation*. Algoritma NBC merupakan pilihan yang lebih baik walau ditambahkan basis pengetahuan eksternal apapun dengan peningkatan akurasi 28,78% dibanding dengan SVM yang memiliki peningkatan akurasi 6,36% saja (Hassan, Rafi, & Shaikh, 2011). Selain algoritma NBC, k-NN merupakan algoritma yang sangat baik untuk penyelesaian masalah klasifikasi walaupun dimasukkan data yang berjumlah sangat besar dibandingkan algoritma lainnya. Pernyataan tersebut didasarkan pada penelitian yang membahas tentang kasus perbandingan algoritma k-NN, NBC dan DT dengan berbagai percobaan *dataset* didalamnya (Jadhav & Channe, 2013). Algoritma k-NN juga sangat baik jika digabungkan dengan seleksi fitur seperti *particle swarm optimization (PSO)* sehingga disebut dengan WPSOk-NN. Bahkan penggabungan kedua algoritma tersebut menghasilkan pengurangan tingkat kesalahan klasifikasi dan meningkatkan efisiensi pada klasifikasi teks berskala besar (Wang & Li, 2010). Walaupun

memiliki tingkat akurasi yang sangat besar, namun pada faktanya k-NN memiliki proses kalkulasi yang lebih lama dibandingkan dengan algoritma seperti NBC dan DT sehingga hal ini diatasi dengan cara memodifikasi algoritma k-NN menjadi Sk-NN. Seperti yang dilakukan oleh Zhang, Huang dan Zhang saat melakukan penelitian tentang kategorisasi teks menggunakan algoritma k-NN. Hasil modifikasi algoritma tersebut dapat menandingi *precision*, *Recall* dan *F1-score* dari algoritma SVM. Bahkan *precision* dan *Recall* memiliki nilai yang sama, perbedaan terletak pada *Recall* yang hanya memiliki selisih 0.01% saja (X. F. Zhang, Huang, & Zhang, 2009). Selain kedua algoritma tersebut, terdapat algoritma lainnya yang juga handal dalam menangani masalah klasifikasi, yaitu algoritma SVM. Algoritma ini populer dalam mengatasi permasalahan klasifikasi berbasis teks. Penelitian yang membahas perbandingan algoritma *supervised learning* menyatakan bahwa SVM merupakan algoritma yang memiliki hasil terbaik dibandingkan dengan algoritma lainnya. Hal ini berdasarkan beberapa atribut seperti akurasi, kecepatan klasifikasi, toleransi terhadap atribut yang tidak relevan, toleransi terhadap data *noise* dan kecepatan belajar (Bhavsar & Ganatra, 2012). Bahkan algoritma SVM berada pada posisi ketiga dari top 10 algoritma terbaik versi IEEE Konferensi Internasional tentang *Data mining* pada Desember 2006 (Wu et al., 2008). Penelitian menggunakan algoritma SVM dilakukan untuk mengklasifikasi teks berjenis buku dengan jumlah 500 judul buku yang terdiri dari 250 judul berkaitan dengan komputer dan 250 judul lainnya berkaitan dengan biologi atau alam (Sarkar, Chatterjee, Das, & Datta, 2015). Walaupun SVM terkenal karena akurasi yang tinggi pada proses klasifikasi teks, namun algoritma ini rentan terhadap *dataset* yang tidak seimbang (*imbalance*) sehingga dibutuhkan penyeimbang *dataset* seperti *SMOTE* (Kwek & Japkowicz, 2004). Selain itu, terdapat penelitian lainnya yang membahas *dataset imbalance* dengan *SMOTE* untuk menaikkan akurasi algoritma SVM (Núñez, Gonzalez-Abril, & Angulo, 2017). Selanjutnya terdapat penelitian yang dilakukan oleh Afyati, Edi dan Anis tentang pengenalan pernyataan sarkas pada WhatsApp Grup berbahasa Indonesia. Metode yang diusulkan pada penelitian ini 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. Hasil akurasi yang didapatkan tidak memadai karena penelitian ini hanya menggunakan data sampling dan eksperimen analisis secara manual (Afiyati, Winarko, & Cherid, 2018).

2.2. Masalah Penelitian

Masalah yang terdapat pada penelitian ini dibagi ke dalam 3 kategori sebagai berikut,

1.2.1. Identifikasi Masalah

WhatsApp merupakan aplikasi olah pesan yang populer di Indonesia. Hal ini menyebabkan munculnya kegiatan serangan baru yang disebut dengan *Cyberbullying*. Penggunaan fitur *end to end encryption* pada WhatsApp membuat pesan obrolan hanya bisa dilihat oleh pengirim dan penerima saja. Penelitian ini akan menggunakan fitur yang disediakan WhatsApp untuk mengeksplor percakapan grup karena *smartphone* yang dipakai dalam keadaan normal.

Cyberbullying menjadi masalah berbahaya karena memiliki dampak yang sangat serius terhadap psikis korban, sehingga pendeteksian *Cyberbullying* sangat penting untuk dilakukan. Pendekatan metode algoritma yang akan digunakan pada klasifikasi tingkah laku *bully* adalah SVM, NBC dan k-NN dengan tambahan seleksi fitur *SMOTE*.

1.2.2. Batasan Masalah

Pembahasan mengenai algoritma yang dapat digunakan untuk proses klasifikasi tingkah laku *bully* sangat luas, sehingga pembahasan pada penelitian ini hanya didasarkan pada komparasi tiga algoritma klasifikasi yaitu SVM, NBC dan k-NN dengan menggunakan *dataset* yang diperoleh dari teks obrolan grup “Pascasarjana 2017” pada aplikasi WhatsApp Messenger. Pelabelan *dataset* tersebut menggunakan metode *crowdsourcing* dengan 21 partisipan. *Field* yang digunakan dalam masalah ini adalah berupa teks dengan tambahan satu atribut

bersifat *label*. Hasil pengolahan data tersebut akan menghasilkan evaluasi akurasi, AUC, *precision*, *recall* dan *F-Measure*. Selain itu, pengujian dengan *T-Test* dilakukan untuk melihat algoritma mana yang lebih baik antara SVM, NBC dan k-NN dalam penyelesaian masalah klasifikasi tingkah laku *bully* pada teks obrolan aplikasi WhatsApp Messenger.

1.2.3. Rumusan Masalah

Beberapa poin rumusan masalah yang didapatkan adalah sebagai berikut,

1. Seberapa tinggi akurasi yang dimiliki algoritma SVM dalam menyelesaikan masalah klasifikasi tingkah laku *bully* pada teks obrolan aplikasi WhatsApp Messenger?
2. Seberapa tinggi akurasi yang dimiliki algoritma NBC dalam menyelesaikan masalah klasifikasi tingkah laku *bully* pada teks obrolan aplikasi WhatsApp Messenger?
3. Seberapa tinggi akurasi yang dimiliki algoritma k-NN dalam menyelesaikan masalah klasifikasi tingkah laku *bully* pada teks obrolan aplikasi WhatsApp Messenger?
4. Metode manakah yang memiliki tingkat akurasi lebih baik antara algoritma SVM, NBC dan k-NN?
5. Apakah ada kenaikan akurasi saat dilakukan penggunaan *SMOTE* pada ketiga algoritma tersebut?
6. Metode manakah yang memiliki tingkat akurasi lebih baik antara algoritma SVM, NBC dan k-NN dengan SVM, NBC dan k-NN ditambah seleksi fitur *SMOTE*?

2.3. Tujuan Penelitian

Tujuan penelitian yang akan dilakukan adalah proses klasifikasi untuk mendeteksi *cyberbullying* menggunakan algoritma SVM, NBC dan k-NN dengan tambahan *SMOTE* berdasarkan data yang sudah disiapkan dan mengkomparasi kinerja algoritma-algoritma tersebut untuk diketahui algoritma mana yang memiliki tingkat akurasi yang lebih baik diantara ketiganya.

1.4. Ruang Lingkup Penelitian

Penelitian ini dibatasi hanya untuk *dataset* obrolan WhatsApp dengan atribut *encrypt12*. Penelitian ini menggunakan fitur yang disediakan WhatsApp untuk mengekspor percakapan grup karena *smartphone* yang dipakai dalam keadaan normal. Selanjutnya dilakukan proses pelabelan pada *dataset* yang diperoleh menggunakan metode *crowdsourcing* dengan 21 partisipan. Setelah itu, proses klasifikasi dilakukan menggunakan metode SVM, NBC dan k-NN. Hasil pengolahan data tersebut akan menghasilkan evaluasi akurasi, AUC, *precision*, *recall* dan *F-Measure*. Selain itu, pengujian dengan *T-Test* dilakukan untuk melihat algoritma mana yang lebih baik antara SVM, NBC dan k-NN dengan tambahan fitur seleksi *SMOTE* dalam penyelesaian masalah klasifikasi tingkah laku *bully* pada teks obrolan aplikasi WhatsApp *Messenger*.

HASIL PENELITIAN

Tabel 4.1. Perbandingan Akurasi dan AUC dari Ketiga Algoritma

Algoritma	Akurasi	AUC
k-NN + <i>SMOTE</i>	78,91%	0,804
NBC + <i>SMOTE</i>	82,59%	0,500
SVM + <i>SMOTE</i>	83,57%	0,925

Tabel 4.27 menunjukkan bahwa akurasi yang dimiliki oleh algoritma SVM + *SMOTE* sebesar 83,57% lebih baik dibandingkan dengan kinerja algoritma NBC + *SMOTE* sebesar 82,59% dan algoritma k-NN + *SMOTE* sebesar 78,91%.

2.4. Evaluasi Algoritma k-NN, NBC dan SVM

a. *Confusion matrix*

Hasil perhitungan akurasi untuk algoritma k-NN, NBC dan SVM dapat dilihat pada tabel 4.28.

Tabel 4.2. *Confusion matrix* Algoritma k-NN, NBC dan SVM

Algoritma	Akurasi	AUC
k-NN	81,32%	0,845
NBC	78,95%	0,500
SVM	81,58%	0,912
k-NN + <i>SMOTE</i>	78,91%	0,804
NBC + <i>SMOTE</i>	82,59%	0,500
SVM + <i>SMOTE</i>	83,57%	0,925

a. *Precision & Recall*

Precision & Recall untuk algoritma k-NN, NBC dan SVM adalah sebagai berikut,

Tabel 4.3. *Precision & Recall* Algoritma k-NN, NBC dan SVM

<i>Algoritma</i>	<i>Precision</i>	<i>Recall</i>
k-NN	82,46%	82,83%
NBC	78,08%	85,38%
SVM	78,55%	90,69%
k-NN + <i>SMOTE</i>	100,00%	57,95%
NBC + <i>SMOTE</i>	80,65%	86,76%
SVM + <i>SMOTE</i>	85,84%	81,26%

b. *F-Measure*

F-Measure untuk algoritma k-NN, NBC dan SVM adalah sebagai berikut,

Tabel 4.4. *F-Measure* Algoritma k-NN, NBC dan SVM

<i>Algoritma</i>	<i>F-Measure</i>
k-NN	82,47%
NBC	81,38%
SVM	84,03%
k-NN + <i>SMOTE</i>	73,01%
NBC + <i>SMOTE</i>	83,28%
SVM + <i>SMOTE</i>	83,08%

Keseluruhan hasil tersebut menunjukkan bahwa akurasi yang dimiliki oleh algoritma SVM + *SMOTE* sebesar 83,57% lebih baik dibandingkan dengan dua algoritma lainnya baik ditambahkan fitur *SMOTE* maupun tidak.

c. *T-Test*

T-Test untuk algoritma k-NN, NBC dan SVM dapat diketahui sebagai berikut,

Tabel 4.5. *T-Test* Algoritma k-NN, NBC dan SVM

A	B	C	D
	0.803 +/- 0.054	0.816 +/- 0.056	0.813 +/- 0.069
0.803 +/- 0.054		0.602	0.709
0.816 +/- 0.056			0.927
0.813 +/- 0.069			

T-Test untuk algoritma k-NN, NBC dan SVM dengan fitur *SMOTE* dapat diketahui sebagai berikut,

Tabel 4.6. *T-Test* Algoritma k-NN, NBC dan SVM dengan fitur *SMOTE*

A	B	C	D
	0.819 +/- 0.045	0.846 +/- 0.042	0.789 +/- 0.047
0.819 +/- 0.045		0.182	0.165
0.846 +/- 0.042			0.011
0.789 +/- 0.047			

Model algoritma yang digunakan dapat memudahkan proses klasifikasi untuk deteksi *bully* dan tidak *bully*. Seluruh data berbentuk kalimat yang diolah akan terpisah menjadi kata-kata setelah dilakukan teks *preprocessing* dan memiliki

bobot pada setiap kata tersebut. Sesuai dengan kelasnya, kata-kata tersebut akan digunakan sebagai data latih untuk mengenali kata-kata lain yang berhubungan dengan klasifikasi *bully*.

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/>
- Alpaydm, E. (2010). *Introduction to Machine Learning Second Edition*.
- Ariawan, I. (2018). Kurva Receiver operating Characteristic (Bagian 1).
- Arwan, Ardina, V., Ariana, L. R., Samuel, F., Ramdani, D., Aditya, & Sukmana, E. A. (2018). Synthetic Minority Over-Sampling Technique (Smote) Algorithm For Handling Imbalanced Data.
- Bhavsar, H., & Ganatra, A. (2012). A Comparative Study of Training Algorithms for Supervised Machine Learning. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(4), 8. <https://doi.org/10.1.1.492.6088>
- 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>
- Connolly, T., & Begg, C. (2010). Database systems : a practical approach to design , implementation , and management, (June 2014).
- Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine Learning*, 1303–1308. https://doi.org/10.1007/978-0-387-73003-5_299
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *Lecture Notes in Computer Science*

- (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 17(3). https://doi.org/10.1007/978-3-319-18032-8_50
- Feldman, R., & Sanger, J. (2007). *The Text Mining HandBook*. <https://doi.org/10.1017/CBO9780511546914>
- Hadi Pratama, A. (2017). Laporan comScore: WhatsApp Adalah Aplikasi Mobile Terpopuler di Indonesia. Retrieved from <https://id.techinasia.com/comscore-whatsapp-adalah-aplikasi-terpopuler-di-indonesia>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. Elsevier Inc.
- Hassan, S., Rafi, M., & Shaikh, M. S. (2011). Comparing SVM and Naïve Bayes classifiers for text categorization with Wikitology as knowledge enrichment. *Proceedings of the 14th IEEE International Multitopic Conference 2011, INMIC 2011*, 31–34. <https://doi.org/10.1109/INMIC.2011.6151495>
- Hinduja, S., & Patchin, J. W. (2005). Cyberbullying Offending, 2005.
- Hoffer, J. A., Ramesh, V., & Topi, H. (2014). *Modern Database Management*.
- Howe, J. (2009). Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business. Jeff Howe. Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business . New York, NY: Crown Business 2008. 320 pp. \$26.95. *Journal of Consumer Marketing*, 26(4), 305–306. <https://doi.org/10.1108/07363760910965918>
- Jadhav, S. D., & Channe, H. P. (2013). Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. *International Journal of Science and Research (IJSR)*, 14611(1), 2319–7064. <https://doi.org/10.1080/00022470.1969.10466496>
- Kothari, C. R. (2004). *Research Methodology: Methods & Techniques*. New Age International (P) Ltd. <https://doi.org/10.1017/CBO9781107415324.004>
- Kwek, R. A. S., & Japkowicz, N. (2004). Applying Support Vector Machines to Imbalanced Datasets. *Proceedings of 15th European Conference on Machine Learning ECML*, 39–50. <https://doi.org/10.1007/978-3-540-30115-8>
- Li, Y., & Jain, A. K. (1998). Classification of Text Documents, 1–3.
- MacLennan, J., Tang, Z., & Crivat, B. (2008). *Data Mining with Microsoft SQL*

- Server 2008. *Animal Genetics* (Vol. 39).
- 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, 6(June), 904–911.
- Núñez, H., Gonzalez-Abril, L., & Angulo, C. (2017). Improving SVM Classification on Imbalanced Datasets by Introducing a New Bias. *Journal of Classification*, 34(3), 427–443. <https://doi.org/10.1007/s00357-017-9242-x>
- Nurhuda, F., Widya Sihwi, S., & Doewes, A. (2016). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Teknologi & Informasi ITSmart*, 2(2), 35. <https://doi.org/10.20961/its.v2i2.630>
- Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques. Advanced Data Mining Techniques*. <https://doi.org/10.1007/978-3-540-76917-0>
- 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>
- Pratiwi, M. D. (2011). Faktor-Faktor yang Mempengaruhi Cyberbullying pada Remaja, 1–11.
- Provost, F., & Fawcett, T. (1997). Analysis and Visualization of Classifier Performance : Comparison under Imprecise Class and Cost Distributions Evaluating and Visualizing Classifier.
- Rahmayani, I. (2015). Indonesia Raksasa Teknologi Digital Asia. Retrieved from https://www.kominfo.go.id/content/detail/6095/indonesia-raksasa-teknologi-digital-asia/0/sorotan_media
- Rasjid, Z. E., & Setiawan, R. (2017). Performance Comparison and Optimization

- of Text Document Classification using k-NN and Naïve Bayes Classification Techniques. *Procedia Computer Science*, 116, 107–112. <https://doi.org/10.1016/j.procs.2017.10.017>
- Sarkar, A., Chatterjee, S., Das, W., & Datta, D. (2015). Text Classification using Support Vector Machine. *International Journal of Engineering Science Invention ISSN*, 4(11), 75–82. <https://doi.org/10.1145/860442.860443>
- Satalina, D. (2014). Kecenderungan Perilaku Cyberbullying Ditinjau Dari Tipe Kepribadian Ekstrovert Dan Introvert, *02(02)*, 294–310.
- Sewell, M. (2009). Kernel Methods, (April 2007), 1–15.
- Turban, E., Aronson, J. E., & Liang, T. (2007). Decision Support Systems and Intelligent Systems, 960.
- UNICEF. (2015). Laporan Tahunan Indonesia 2015. *UNICEF Laporan Tahunan Indonesia 2015*, 1–19. Retrieved from https://www.unicef.org/indonesia/id/Laporan_Tahunan_UNICEF_Indonesia_2015.pdf
- Vapnik, V. (1999). *The Nature of Statistical Learning Springer* (Second Edi). Springer-Verlag: New York Berlin Heidelberg.
- Vercellis, C. (2009). *Business Intelligence*.
- Vijayarani, S., Ilamathi, J., Nithya, M., Ilamathi, M. J., Nithya, M., Professor, A., & Research Scholar, M. P. (2015). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16. <https://doi.org/10.1016/j.procs.2013.05.286>
- W Rahmawati, S. (2016). Salah Kaprah Istilah Bullying. Retrieved from <http://buletin.k-pin.org/index.php/arsip-artikel/45-salah-kaprah-istilah-bullying>
- Wang, L., & Li, X. (2010). An improved KNN algorithm for text classification, 436–439.
- WhatsApp Inc. (2018). Enkripsi end-to-end. Retrieved from <https://faq.whatsapp.com/en/android/28030015/?lang=id>
- Wicaksana, I. W. S., C, D. A., Baskoro, D. A., & Ambarwati, L. (2013). *Belajar Data Mining dengan RapidMiner*.

- Willard, N. (2007). Educator ' s Guide to Cyberbullying and Cyberthreats, 1–16.
- Williams, G. J., & Huang, Z. (1996). Modelling the KDD Process, (June), 1–8.
- Windu Gata. (2018). Text Mining Program. Retrieved from <http://www.gataframework.com/textmining>
- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). *Top 10 algorithms in data mining. Knowledge and Information Systems* (Vol. 14). <https://doi.org/10.1007/s10115-007-0114-2>
- Zhang, H., & Li, D. (2007). Naive Bayes text classifier. *Proceedings - 2007 IEEE International Conference on Granular Computing, GrC 2007*, (3), 708–711. <https://doi.org/10.1109/GRC.2007.4403192>
- Zhang, X. F., Huang, H. Y., & Zhang, K. L. (2009). KNN text categorization algorithm based on semantic centre. *Proceedings - 2009 International Conference on Information Technology and Computer Science, ITCS 2009*, 1, 249–252. <https://doi.org/10.1109/ITCS.2009.57>

