

KOMPARASI OPTIMASI ANALISIS SENTIMEN CYBERBULLYING PADA INSTAGRAM BERBASIS PARTICLE SWARM OPTIMIZATION

Asti Herliana¹, Shofiyah Siti Muawiyah²

¹Universitas Adhirajasa Reswara Sanjaya
e-mail korespondensi: asti@ars.ac.id

²Universitas Adhirajasa Reswara Sanjaya
e-mail: shfhsiti@gmail.com

Abstrak

Sejak masa pandemi covid-19 melanda dunia, sekitar 78,19% manusia di Indonesia mengandalkan media internet sebagai penunjang utama kegiatan sehari-hari. Hal ini membuat aktivitas manusia mayoritas dilakukan melalui dunia maya, salah satunya adalah sebagai bentuk eksistensi. Media sosial seperti Instagram, menjadi pilihan dari banyak manusia di dunia utamanya Indonesia untuk menyalurkan segala macam bentuk aspirasinya. Efek dari meningkatnya postingan di media sosial ini juga berimbas kepada tingginya tingkat perundungan melalui dunia maya yang sering dikenal dengan istilah *cyberbullying*. Salah satu bentuk *cyberbullying* yang marak terjadi adalah melalui ujaran kebencian dan kata-kata yang tidak baik terhadap postingan yang diunggah. Pada penelitian kali ini akan dilakukan optimasi untuk mengetahui analisis sentimen terhadap berbagai tindak *cyberbullying* yang ada pada media sosial Instagram agar dapat ditindak lebih lanjut dengan menggunakan metode *Particle Swarm Optimization* (PSO). Sedangkan untuk metode klasifikasi dari analisis sentiment pada penelitian kali ini dilakukan komparasi dengan menggunakan algoritma *support vector machine* (SVM) dan *naïve bayes*. Dari hasil penelitian diketahui bahwa performa metode PSO memberikan hasil yang lebih baik jika dikombinasikan dengan metode SVM yang mencapai nilai akurasi 78,60% dengan dukungan 100% *class precission*. Sedangkan hasil *naïve bayes* hanya mencapai nilai akurasi 78,00% dengan dukungan *class precission* sebesar 99,74%.

Kata Kunci: Komparasi, Optimasi, Analisis Sentimen, *Cyberbullying*, PSO, *Naïve Bayes*, Support Vector Machine.

Abstract

Since the Covid-19 pandemic hit the world, almost 78,19% of people in Indonesia have relied on internet media as the main support for their daily activities. This means that the majority of human activities are carried out through cyberspace, one of which is as a form of existence. Social media, such as Instagram, is the choice of many people in the world, especially Indonesia, to channel all kinds of aspirations. The effect of increasing posts on social media has also resulted in high levels of bullying through cyberspace, which is often known as cyberbullying. One form of cyberbullying that is widespread is through hate speech and unkind words regarding uploaded posts. In this research, optimization will be carried out to determine sentiment analysis of various acts of cyberbullying on Instagram social media so that further action can be taken using the Particle Swarm Optimization (PSO) method. Meanwhile, the classification method for sentiment analysis in this research was compared using the support vector machine (SVM) and naïve Bayes algorithms. From the research results, it is known that the performance of the PSO method provides better results when combined with the SVM method which reaches an accuracy value of 78.60% with support for 100% class precision. Meanwhile, the Naïve Bayes results only reached an accuracy value of 78.00% with class precision support of 99.74%.

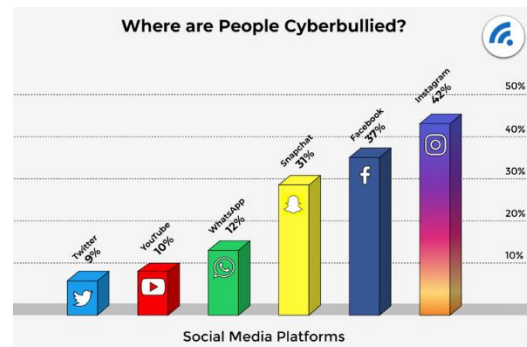
Keywords: Comparison, Optimization, Sentiment Analysis, *Cyberbullying*, PSO, *Naïve Bayes*.

1. Pendahuluan

Semenjak pandemi Covid-19 melanda dunia, pola kegiatan Masyarakat dunia banyak mengalami perubahan. Hal ini juga mengubah paradigma sosial manusia menjadi lebih meningkat di dunia maya. Hal ini ditunjukkan dengan adanya survei bahwa pengguna media sosial Instagram meningkat sebesar 5,01% pada bulan Desember 2022 menjadi sebanyak 92,53 juta pengguna (Mustajab, 2021). Dengan meningkatnya pengguna media sosial seperti Instagram, maka terdapat beberapa dampak negative yang ditimbulkan. Salah satu bentuk dampak yang ditimbulkan adalah maraknya kejahatan seperti *cyberbullying*. Seperti yang dilansir oleh *Center for Digital Society (CfDS)* pada bulan Agustus 2021, bahwa hasil penelitiannya yang dilakukan terhadap remaja usia 13-18 tahun di 34 provinsi yang berada di Indonesia tercatat sekitar 1.895 siswa telah menjadi korban *cyberbullying* (Fahlevi & Sutriyanto, 2023).

Cyberbullying dapat diartikan sebagai bentuk penyalahgunaan yang dimaksudkan untuk melecehkan, mengancam, mempermalukan dan mengejek orang lain (Destitus, Wella, & Suryasari, 2020). Pada dasarnya kasus *cyberbullying* melalui platform media sosial seperti Instagram merupakan salah satu upaya untuk mengintimidasi dan menggretak orang dengan mengirim pesan mengintimidasi atau mengancam, *Cyberbullying* juga menjadikan ujaran kebencian sebagai salah satu bentuk intimidasi terhadap seseorang atau kelompok tertentu (Rachmayanti & Candrasari, 2022).

Tingginya kasus *cyberbullying* khususnya di Indonesia salah satu penyebabnya adalah adanya peran media sosial. Dan menurut data statistik menunjukkan fakta bahwa media sosial Instagram merupakan platform dengan kasus *cyberbullying* tertinggi. Fakta mengenai hal ini ditunjukkan dengan adanya grafik hasil survey kasus *cyberbullying* sebagai berikut :



Gambar 1. Presentase Media Sosial Paling Banyak Menimbulkan *Cyberbullying*
Sumber : (Pratama & Nistanto, 2021)

Berdasarkan data pada gambar 1 dapat disimpulkan bahwa media sosial utamanya Instagram memegang peranan penting akan tingginya kasus *cyberbullying* khususnya di Indonesia.

Tingginya kasus *cyberbullying* di Indonesia, memerlukan perhatian khusus agar dapat dilakukan penanggulangan lebih lanjut. Salah satu cara yang dapat dilakukan untuk mengetahui respon terhadap kasus ini adalah dengan melakukan analisis sentiment berbasis *text mining*. Analisis sentiment menurut Tetsuya Nasukawa dan Jeonghee Yi dapat diartikan sebagai sebuah teknik yang digunakan untuk mengidentifikasi teks untuk kemudian dikategorikan kedalam sentiment positif atau sentiment negatif (Maulana & Ernawati, 2020). Sedangkan *text mining* dapat diistilahkan sebagai penambangan teks dimana para ahli mentafsirkannya sebagai proses penambangan data untuk memperoleh informasi berkualitas tinggi dari suatu teks (Deolika, Kusriani, & Luthfi, 2019). Selain itu, *text mining* juga dapat diartikan sebagai sebuah teknik algoritmik yang berbasis komputer dengan tujuan untuk mendapatkan pengetahuan baru yang tersembunyi dari sekumpulan teks (Priyanto & Ma'arif, 2018).

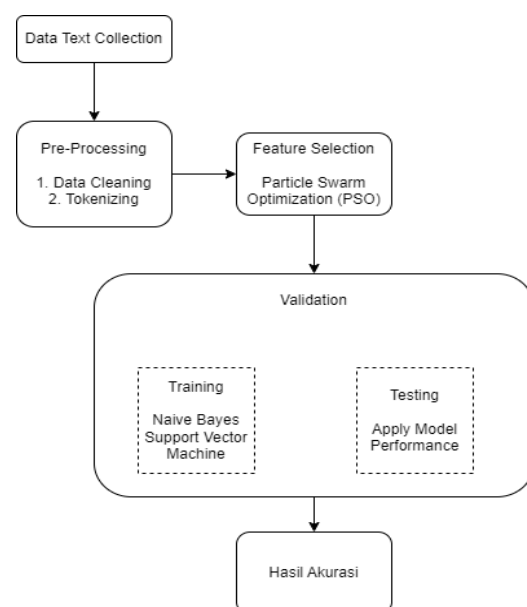
Penelitian mengenai analisis sentiment sendiri telah banyak dilakukan pada contoh permasalahan yang serupa seperti yang dibahas pada penelitian kali ini. Adapun beberapa penelitian mengenai analisis sentiment sebelumnya pernah diterapkan untuk mengetahui respon Masyarakat terhadap aplikasi investasi saham (Rizkiansyah, Herliana, Alamsyah, & Tjoe, 2022). Selain itu, analisis sentiment juga dilakukan untuk mengetahui respon Masyarakat terhadap adanya perkuliahan secara daring selama pandemic Covid-19 (Ernamia & Herliana, 2022).

Selain penelitian mengenai analisis sentiment, untuk solusi permasalahan berupa *text mining* dengan menggunakan algoritma *support vector machine* (SVM) dan *naïve bayes* pada penelitian kali ini juga telah diajukan oleh beberapa peneliti terdahulu. Adapun beberapa penelitian yang dijadikan rujukan solusi penyelesaian masalah pada penelitian kali ini adalah penelitian yang dilakukan oleh Fanny, Yohan Muliono dan Fidelson Tanzil pada tahun 2018 dimana pada penelitian ini dilakukan komparasi metode K-NN, *naïve bayes* dan SVM untuk mengkategorisasikan berita melalui media internet (Fanny, Muliono, & Tanzil, 2018). Dari penelitian pada tahun 2018 ini, didapatkan hasil berupa kategorisasi berita melalui media internet dengan metode K-NN dan SVM sebagai metode terbaik dengan hasil performansi akurasi diatas 80%. Penelitian selanjutnya yang juga menggunakan metode SVM dan *naïve bayes* terhadap konteks *text mining* adalah penelitian yang dilakukan oleh Ardianto, dkk pada tahun 2020 dimana pada penelitian ini dilakukan pengklasifikasian *text mining* untuk analisis sentiemen terhadap Pendidikan e-sport melalui media sosial twitter. Pada penelitian ini algoritma *naïve bayes* menunjukkan hasil performansi lebih baik dengan mencapai hasil akurasi sebesar 70,32% (Ardianto, Rivanie, Alkhalifi, Nugraha, & Gata, 2020). Penelitian terakhir yang dijadikan rujukan solusi permasalahan dari penelitian kali ini adalah penelitian yang dilakukan oleh Muhammad, Bukhori dan Pandunata pada tahun 2019 dimana pada penelitian ini dilakukan proses klasifikasi untuk mengetahui analisis sentiment terhadap konten pada channel youtube. Dengan pendekatan *text mining*, pada penelitian ini didapatkan performansi presisi terbaik sebesar 91% terhadap kombinasi metode SVM dengan *naïve bayes* (Muhammad, Bukhori, & Pandunata, 2019). Pada penelitian yang dilakukan pada kesempatan ini, dilakukan komparasi dari beberapa metode klasifikasi seperti *naïve bayes* dan SVM yang telah memberikan hasil persentase terbaik untuk jenis data yang serupa dengan penelitian yang dilakukan pada penelitian kali ini untuk dikombinasikan dengan metode PSO. Hal ini dilakukan untuk mengetahui metode mana yang memberikan hasil kombinasi terbaik sehingga dapat dijadikan acuan pada penelitian selanjutnya dengan data yang berbeda.

2. Metode Penelitian

Metode penelitian yang digunakan pada penelitian kali ini adalah penelitian kualitatif. Hal ini sesuai dengan salah satu karakteristik dari penelitian kualitatif sendiri dimana pengumpulan data dilakukan atas dasar prinsip fenomenologis (Tysara, 2023). Metode kualitatif dapat diartikan sebagai sebuah metode penelitian yang digunakan pada kondisi ilmiah Dimana peneliti adalah sebagai instrument, teknik pengumpulan data dan dianalisis yang menekankan pada makna (Sugiyono, 2018). Untuk menjabarkan Langkah-langkah yang dilakukan pada penelitian kali ini, berikut ini disampaikan gambarannya melalui metodologi penelitian.

Metodologi Penelitian



Gambar 2. Metodologi Penelitian

Dari gambar 2 dapat dijelaskan Langkah-langkah penelitian sebagai berikut :

1. Pada penelitian kali ini digunakan data awal berupa data text yang didapatkan dari kolom komentar pada media sosial Instagram dan bersifat kalimat *bullying*. Pengambilan data text dilakukan dengan memanfaatkan aplikasi API (*Application Program Interface*).
2. Setelah data text yang terdapat kalimat *bullying* didalamnya didapatkan, selanjutnya data text di berikan perlakuan tahap *pre-processing* berupa *cleaning* dan *tokenizing*. Pada proses *cleaning*, data yang didapatkan diberikan perlakuan 4 proses pembersihan yakni dengan menggunakan operator *select attribute*

- (menghilangkan atribut yang tidak digunakan dalam analisis sentimen), *remove duplicates* (menghilangkan data ganda, pada penelitian ini semula data berjumlah 500 data, berkurang menjadi 496 data), *replace missing value* (menghilangkan data yang tidak memiliki nilai untuk penelitian) dan *Subprocess* (digunakan untuk menghilangkan data yang bersifat simbol dan emoji, sehingga data yang dihasilkan hanya data berupa kata saja).
3. Tahapan pre-processing selanjutnya yang dilakukan adalah *tokenizing*. Pada proses ini dilakukan pemisahan kata pada setiap teks yang ada. Terdapat beberapa proses yang dilakukan untuk tahap *tokenizing* ini. Tahap pertama yakni penambahan atribut *set role* (pembeda baris penamaan atribut koordinat dan posisi yang akan dimasukkan kedalam kategori label) yang dilanjutkan dengan penambahan atribut *nominal to text* (berfungsi menjadikan data nominal menjadi teks). Setelah penambahan atribut *nominal to text*, Langkah selanjutnya dari proses *tokenizing* ini ditambahkan atribut *process document from data* yang berfungsi sebagai tempat penyimpanan proses *tokenizing*.
 4. Setelah data melalui tahap pre-processing dimana data yang ditambah telah bersih dan siap untuk diproses lebih lanjut, data tersebut selanjutnya dilakukan optimasi dengan menggunakan metode *particle swarm optimization* (PSO) yang dilanjutkan dengan klasifikasi dengan metode naïve bayes dan *support vector machine* (SVM). Kedua metode klasifikasi ini dikomparasi untuk mendapatkan hasil metode klasifikasi mana yang terbaik jika dikombinasikan dengan metode optimasi PSO. Pada proses klasifikasi data yang didapat dari hasil pre-processing, akan dibagi menjadi data latih dan data uji untuk kemudian akan dikelompokkan menjadi hasil kelas positif dan hasil kelas negative.

Pada penelitian kali ini terdapat beberapa landasan teori yang dijadikan dasar dan fakta. Adapun beberapa landasan teori yang dimaksud adalah :

1. Support Vector Machine
 Pada penelitian kali ini terdapat 2 metode yang dilakukan komparasi yakni *Support Vector Machine* (SVM) dan Naïve Bayes. SVM sendiri secara sederhana dapat diartikan sebagai usaha untuk mencari

hyperplane terbaik yang berperan untuk mencari garis batas dari dua buah *class* (Jumeilah, 2017). Kinerja dari algoritma SVM didasarkan pada *Structural Risk Minimization* (SRM) yang dirancang untuk mengolah data menjadi *hyperplane* yang mengklasifikasikan ruang *input* menjadi dua kelas.

2. Naïve Bayes
 Algoritma kedua yang dijadikan komparasi dan dikombinasikan dengan algoritma *Particle Swarm Optimization* (PSO) adalah naïve bayes. Naïve bayes adalah salah satu algoritma pengklasifikasian Dimana pembelajaran ditekankan pada pengestimasian probabilitas. Metode pengklasifikasian dengan metode probabilitas dan statistic ini pertama kali dikemukakan oleh Thomas Bayes (Watratan, Puspita, & Moeis, 2020). Salah satu manfaat dari algoritma naïve bayes yang dimanfaatkan pada penelitian kali ini adalah manfaat keberhasilan dalam proses *text classification*.
3. Particle Swarm Optimization
 Pada penelitian kali ini, yang dijadikan metode optimasi utama adalah *Particle Swarm Optimization* (PSO) algoritma ini dipilih karena memiliki sifat mudah diimplementasikan, efisien secara komputasi, fleksibel dan mudah untuk diadaptasi pada permasalahan optimasi yang berbeda. PSO sendiri merupakan algoritma metaheuristic (tidak terikat pada aturan atau struktur masalah tertentu) yang digunakan untuk mengatasi permasalahan optimasi yang terinspirasi dari perilaku kawanan hewan seperti burung atau ikan dalam mencari makan (Pajri, Umaidah, & Padilah, 2020).

3. Hasil dan Pembahasan

Penerapan metode optimasi *Particle Swarm Optimization* terhadap metode naïve bayes dan *Support Vector Machine* (SVM) pada penelitian kali ini dilakukan untuk menemukan metode kombinasi yang terbaik untuk analisis sentiment yang diperoleh dari kolom komentar masyarakat pada media sosial Instagram.

3.1. Hasil Pengumpulan Data

Proses pengumpulan data komentar Instagram ini menggunakan pencarian melalui hastag yang sedang trending pada media sosial yang kemudian dikumpulkan kedalam microsoft Excel sebelum data diolah menggunakan RapidMiner.

NO	UserName	At	Content	sumber
1	renniku	22/06/2023	kaya, keluarga baik2, suami ganteng. Kurang apalgi??? Kurang bers	cumicumi.com_insta
2	apa_aja103	23/06/2023	kurang ajar	cumicumi.com_insta
3	tori.barcelonasalau	23/06/2023 18:00	gak ngurus suami mlh ngurusnya selingkuh, istri kurang bersukur	cumicumi.com_insta
4	donnalimbong	25/06/2023 09:56	kerja gaji dapat puluhan juta ngelul! Noh yang ibu2 dagang ampe ten	diadona.id
5	mrs.dianfine_sitorus	25/06/2023 11:59	Ratu drama	diadona.id
6	kanasahid	25/06/2023 14:00	Alay	diadona.id
7	ummi9509	25/06/2023 15:00	cieeee... emang bener kata pepatah gajah di pelupuk mata kaga keliz	indonesia_anies2022
8	deddyhardinal	25/06/2023 15:03	beresin tuh daerah anda.... Banjir dan jalan rusak belum lagi tingkat k	indonesia_anies2022
9	pojok_setingdesain	25/06/2023 15:06	kebanyakan drama lu jar..	indonesia_anies2022
10	ghina.putri123	25/06/2023 15:59	mau ketawa takut dikira ngehujat, padahal oon-nya udah keliatan	indonesia_anies2022
11	mardijonchaniago	25/06/2023 16:12	kayanya makin stres bapak ini	indonesia_anies2022
12	badawi_bdw	25/06/2023 07:34	kocak udah berasa presiden kali dia, gak tau diri apa gk tau malu ya	negeribegundal2.0
13	fadhullohfaishol	25/06/2023 07:35	pencitraannya jauh lebih ekstrim nih org	negeribegundal2.0
14	ade_zahra_batoebar	25/06/2023 07:38	mau muntah liat boneka boneka ni.	negeribegundal2.0
15	renojak36	25/06/2023 07:39	pencitraan yg bagong daerah sendiri gak di urus	negeribegundal2.0

Gambar 3. Proses Pengumpulan Data Secara Manual Dengan Excel

3.2. Hasil Pembacaan Data

Proses pembacaan data atau proses *Read excel* data komentar Instagram ini menggunakan operator *Read Excel*, yaitu suatu layanan yang berisikan suatu perintah, fungsi dan komponen untuk membaca data yang akan dimasukkan. *Tool* operator yang digunakan untuk proses pengambilan data yang akan diolah dengan menggunakan RapidMiner.

Pemilihan Rapid Miner sebagai pengolah analisis sentimen dikarenakan RapidMiner merupakan tools pemograman yang umum di gunakan sebagai pengolahan data Analisis Sentimen. Setelah proses *read excel* data berhasil dilakukan, didapatkan data mentah seperti Gambar 4 berikut ini.

Row No.	NO	UserName	At	Content	sumber
1	1	renniku	Jun 22, 2023 9:00:00 PM ICT	kaya, keluarga baik2, suami ganteng. Kurang apalgi??? Kurang bersukur	cumicumi.com
2	2	apa_aja103	Jun 23, 2023 7:00:00 PM ICT	kurang ajar	cumicumi.com
3	3	tori.barcelonasalau	Jun 23, 2023 9:00:00 PM ICT	gak ngurus suami mlh ngurusnya selingkuh, istri kurang bersukur	cumicumi.com
4	4	donnalimbong	Jun 25, 2023 9:56:00 AM ICT	kerja gaji dapat puluhan juta ngelul! Noh yang ibu2 dagang ampe ten	diadona.id
5	5	mrs.dianfine_sitorus	Jun 25, 2023 11:59:00 AM ICT	Ratu drama	diadona.id
6	6	kanasahid	Jun 25, 2023 7:00:00 PM ICT	Alay	diadona.id
7	7	ummi9509	Jun 25, 2023 3:00:00 PM ICT	cieeee... emang bener kata pepatah gajah di pelupuk mata kaga keliz	indonesia_anies2022
8	8	deddyhardinal	Jun 25, 2023 3:03:00 PM ICT	beresin tuh daerah anda.... Banjir dan jalan rusak belum lagi tingkat k	indonesia_anies2022
9	9	pojok_setingdesain	Jun 25, 2023 3:06:00 PM ICT	kebanyakan drama lu jar..	indonesia_anies2022
10	10	ghina.putri123	Jun 25, 2023 3:59:00 PM ICT	mau ketawa takut dikira ngehujat, padahal oon-nya udah keliatan	indonesia_anies2022
11	11	mardijonchaniago	Jun 25, 2023 4:12:00 PM ICT	kayanya makin stres bapak ini	indonesia_anies2022
12	12	badawi_bdw	Jun 25, 2023 7:34:00 AM ICT	kocak udah berasa presiden kali dia, gak tau diri apa gk tau malu ya	negeribegundal2.0
13	13	fadhullohfaishol	Jun 25, 2023 7:35:00 AM ICT	pencitraannya jauh lebih ekstrim nih org	negeribegundal2.0
14	14	ade_zahra_batoebar	Jun 25, 2023 7:38:00 AM ICT	mau muntah liat boneka boneka ni.	negeribegundal2.0
15	15	renojak36	Jun 25, 2023 7:39:00 AM ICT	pencitraan yg bagong daerah sendiri gak di urus	negeribegundal2.0

Gambar 4. Data Keluaran read excel menggunakan RapidMiner

3.3. Hasil Pre-Processing

Data komentar yang diperoleh dari kolom komentar Instagram masih berupa data mentah sehingga di perlukan

preprocessing data untuk memperoleh data yang bersih dan terstruktur supaya dapat digunakan untuk klasifikasi sentimen. Pada saat *preprocessing* data terdiri dari beberapa

tahapan. Terdapat 2 tahap dalam proses *preprocessing* untuk memperoleh hasil yang maksimal. Adapun tahapan sebagai berikut:

1. Proses *Cleaning*

Langkah yang menjadi bagian dalam proses *cleaning* data diantaranya :

- Select Attribute*, pada proses ini akan menghilangkan atribut yang tidak digunakan dalam analisis sentimen.
- Remove duplicates* pada penelitian ini data awal yang diperoleh dari proses sebelumnya berjumlah 500 data, setelah dilakukan proses *remove duplicates* maka tersisa 496 data komentar yang akan digunakan pada tahap selanjutnya
- Replace Missing Value*, dalam proses ini, data yang telah diperoleh dicari data yang tidak bernilai atau menghilangkan data yang tidak bernilai. Dalam penelitian ini data yang diperoleh tidak terdapat *Mising Value* atau data yang tidak bernilai pada komentar.
- Subprocess* dengan memasukan *Replace* sebagai bagian dari proses *subprocess*, pada proses ini *subprocess* digunakan sebagai tempat atau wadah untuk menampung *replace*. Terdapat 2 *replace* yang dipilih dalam subproses, dalam masing-masing *replace* terdapat proses untuk menghilangkan data yang tidak diperlukan. Setelah proses *cleaning* selesai, maka didapat data yang sudah bersih dan dapat digunakan pada proses selanjutnya.

Dari serangkaian proses *cleaning* didapatkan hasil *pre-processing* sebagai berikut :

Tabel 1. Hasil *Pre-Processing Data Cleaning*

DATA CRAWLING	DATA CLEANING
Gaji serius , kerja bercanda. Top 😊	Gaji serius kerja bercanda Top
Bangsat !!	Bangsat
Gokil coy nge slot pake uang rakyat 😊	Gokil coy nge slot pake uang rakyat
Ya gmana mau di Brantas atau di Blokir Judi Onlen toh mereka juga doyan... Adiknya pdhl baru masuk islam si teteh malah leps hijab ya salam...	Ya gmana mau di Brantas atau di Blokir Judi Onlen toh mereka juga doyan Adiknya pdhl baru masuk islam si teteh malah leps hijab ya salam

🤔🤔🤔🤔🤔 it zeusss ngapa ke candy Crush. Udah ketauan bukan nya maaf malah nyari alasan 😊	it zeusss ngapa ke candy Crush Udah ketauan bukan nya maaf malah nyari alasan
Toleransi sesama agama 😊	Toleransi sesama agama
Namanya juga wakil rakyat..itu mewakili rakyat buat main game... 😊😊😊	Namanya juga wakil rakyatitu mewakili rakyat buat main game
Di samping kanan kiri pak Jokowi ada yg korupsi kah???	Di samping kanan kiri pak Jokowi ada yg korupsi kah
polisi kalau mau nilang. Minta tolong tunjukan sertifikatnya..... 😊😊😊😊😊	polisi kalau mau nilang Minta tolong tunjukan sertifikatnya
Dari polisi sampai pelajar doyan mabukhancur sudah !!!!!	Dari polisi sampai pelajar doyan mabuk hancur sudah

2. Proses Tokenizing

Tahap *pre-processing* selanjutnya adalah proses tokenizing. Pada proses ini terdapat beberapa proses meliputi yang pertama penambahan atribut *Set Role*, atribut ini berfungsi sebagai pembeda baris penamaan atribut koordinat dan posisi yang akan dimasukkan kedalam kategori label. Yang kedua penambahan atribut *Nominal to Text*, atribut ini berfungsi sebagai data nominal menjadi teks, yang ketiga yaitu penambahan atribut *process document from data*, atribut ini berfungsi sebagai tempat menyimpan proses *tokenizing* dan yang keempat penambahan proses *Cross Validation* yang memiliki dua subproses, yaitu subproses training yang digunakan untuk melatih model dan subproses testing untuk pengujian sekaligus mengukur kinerja model.

Adapun 2 Operator yang dijalankan didalam proses *preprocessing*, yang pertama *Tokenize* yaitu pemecahan *string* kalimat menjadi kata, yang kedua yaitu operator *Transform cases*, elemen ini berfungsi sebagai pengubahan data menjadi huruf kecil. Dari proses tokenizing, didapatkan data komentar sebagai berikut :

Row No.	word	in documents	total
1	udah	18	18
2	agama	15	16
3	hidup	15	16
4	cantik	10	14
5	kalo	12	13
6	rakyat	12	13
7	anaknya	8	10
8	duit	10	10
9	kaya	9	10
10	pake	8	10
11	sosial	9	10
12	allah	8	9
13	banget	8	8
14	emang	8	8
15	korupsi	8	8
16	negara	8	8
17	salah	8	8
18	video	7	8

Gambar 5. Wordlist Tokenizing Klasifikasi

Berdasarkan hasil keseluruhan proses *tokenizing*, didapatkan hasil sebagai berikut :

Tabel 2. Hasil *Tokenizing*

Sebelum Proses	Sesudah Proses
Gaji serius kerja bercanda Top	'gaji' 'serius' 'kerja' 'bercanda' 'top'
Bangsat	'bangsat'
Gokil coy nge slot pake uang rakyat	'gokil' 'cuy' 'nge' 'slot' 'pake' 'uang' 'rakyat'
Ya gmana mau di Brantas atau di Blokir Judi Onlen toh mereka juga doyan	'ya' 'gmna' 'mau' 'di' 'brantas' 'atay' 'di' 'blokir' 'judi' 'onlen' 'toh' 'mereka' 'juga' 'doyan'
Adiknya pdhl baru masuk islam si teteh malah leps hijab ya salam	'adiknya' 'pdhl' 'baru' 'masuk' 'islam' 'si' 'teteh' 'malah' 'lepas' 'hijab' 'ya' 'salam'
it zeussss ngapa ke candy Crush Udah ketauan bukan nya maaf malah nyari alas an	'it' 'zeussss' 'ngapa' 'ke' 'candy' 'crush' 'udah' 'ketauan' 'bukannya' 'maaf' 'malah' 'nyari' 'alasan'

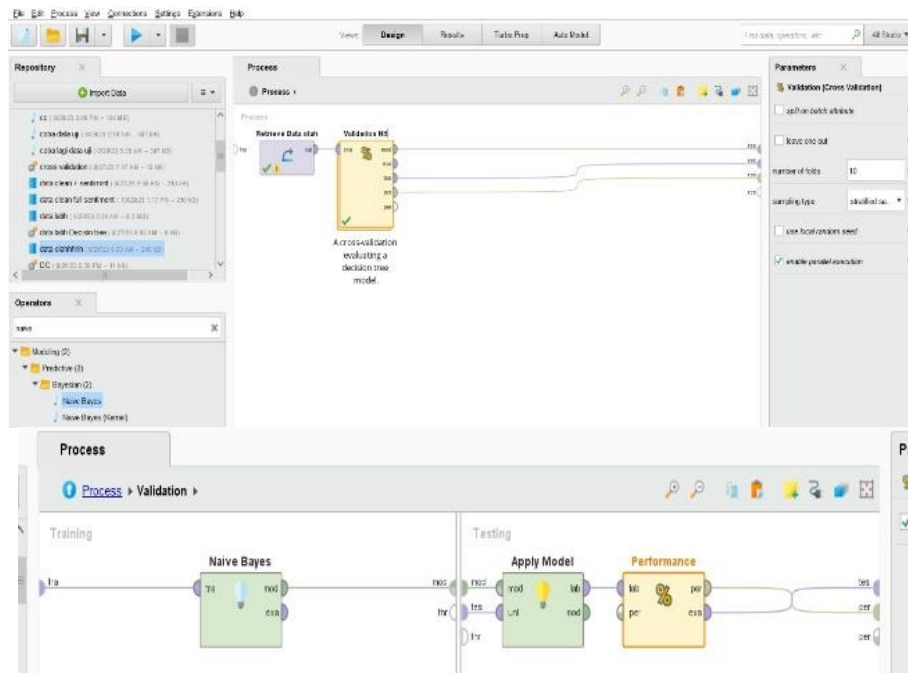
Toleransi sesama agama	'toleransi' 'sesama' 'agama'
Namanya juga wakil rakyatitu mewakili rakyat buat main game	'namanya' 'juga' 'wakil' 'rakyat' 'itu' 'mewakili' 'rakyat' 'buat' 'main' 'game'
Di samping kanan kiri pak Jokowi ada yg korupsi kah	'di' 'samping' 'kanan' 'kiri' 'pak' 'jokowi' 'ada' 'yg' 'korupsi' 'kah'
polisi kalau mau nilang Minta tolong tunjukan sertifikatnya	'polisi' 'kalau' 'mau' 'nilang' 'minta' 'tolong' 'tunjukin' 'sertifikatnya'
Dari polisi sampai pelajar doyan mabuk hancur sudah	'dari' 'polisi' 'sampai' 'pelajar' 'doyan' 'mabuk' 'hancur' 'sudah'

3.4. Hasil Pengujian

Seperti yang telah dibahas pada tahap metodologi penelitian, pada hasil pengujian akan dibandingkan antara hasil optimasi dan klasifikasi antar PSO dan naïve bayes dengan hasil klasifikasi antara PSO dan SVM. Adapun hasil pengujian disampaikan sebagai berikut :

3.4.1. Naïve Bayes dan PSO

Berdasarkan dari penelitian, didapatkan model pengujian naïve bayes dan PSO sebagai berikut :



Gambar 6. Model Proses Optimasi PSO dan Klasifikasi Naïve Bayes.

Berdasarkan dari model yang didapat, maka hasil proses dari optimasi PSO terhadap metode naïve bayes pada analisis

sentiment *cyberbullying* pada media Instagram didapatkan hasil *example set* sebagai berikut :

The screenshot shows the 'ExampleSet (Retrieve data clean full sentiment)' window in Orange Data Mining. It displays a table with 10 rows of data. The first row is POSITIF, and the remaining 9 rows are NEGATIF. The content of the rows is as follows:

Row No.	Sentiment	Content
1	POSITIF	kayakelarga baik2 suami ganteng Kurang...
2	NEGATIF	kurang ajar
3	NEGATIF	gak ngurus suami mih ngurusya selingku.
4	NEGATIF	kenja gaji cepat pulhan juta ngeku Mbh yan.
5	NEGATIF	Ratu drama
6	NEGATIF	Aley
7	NEGATIF	cieee emang bener kaka pepatah gaah di p.
8	NEGATIF	beresin tuh daerah anda Banyir dan jalan nu.
9	NEGATIF	kebanyakan drama lu jar
10	NEGATIF	mau ketawa takut dikira ngehujat pedanal...

Gambar 7. Hasil *Example Set* data metode PSO+Naïve Bayes

Berdasarkan hasil klasifikasi model PSO+Naïve bayes, didapatkan hasil lain berupa akurasi performansi model sebesar

78,00%. Hasil lengkap dari performansi model PSO dan naïve bayes disampaikan sebagai berikut :

	true POSITIF	true NEGATIF	class precision
pred. POSITIF	0	1	0.00%
pred. NEGATIF	109	390	78.16%
class recall	0.00%	99.74%	

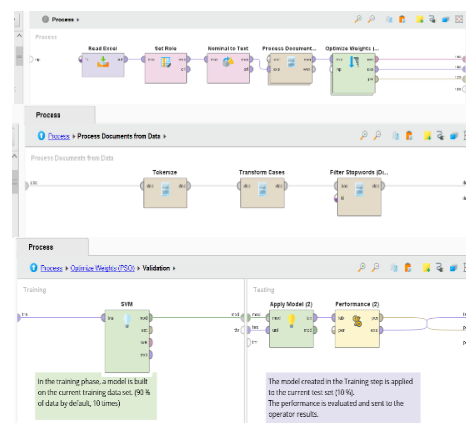
accuracy: 78.00% +/- 0.94% (micro average: 78.00%)

Gambar 8. Hasil Performansi Metode PSO+Naïve Bayes.

Dari performansi yang dihasilkan, maka dapat diketahui bahwa untuk class precision prediksi positif adalah 0.00%, untuk class precision prediksi negatif sebesar 78.16% dan recall yang diperoleh dari true positif sebesar 0.00% sedangkan untuk true negatif sebesar 99,74%.

3.4.2. SVM dan PSO

Selain model dengan menggunakan metode PSO+naïve bayes, pada penelitian kali ini juga dilakukan pengujian untuk model klasifikasi dengan optimasi dimana metode PSO ditambahkan untuk menguji akurasi klasifikasi terhadap metode SVM. Adapun model yang dihasilkan dengan menggunakan rapid miner untuk pengujian PSO+SVM disampaikan sebagai berikut :



Gambar 9. Model Proses Optimasi PSO dan Klasifikasi SVM

Hasil pertama yang didapatkan dari model proses optimasi PSO dan klasifikasi SVM adalah berupa hasil *example set*. Adapun hasil yang dimaksud disampaikan sebagai berikut :

Row No.	Sentiment	Content
1	POSITIF	kayakeluarga baik2 suami ganteng Kurang..
2	NEGATIF	kurang ajar
3	NEGATIF	gak ngurus suami mih ngurusnya selingku..
4	NEGATIF	kerja gaji dapat puluhan juta ngelu Noh yan.
5	NEGATIF	Ratu drama
6	NEGATIF	Alay
7	NEGATIF	cieeee emang bener kata pepatah gajah di p..
8	NEGATIF	beresin tuhdaerah anda Banjir dan jalan ru..
9	NEGATIF	kebanyakan drama lu jar
10	NEGATIF	mau ketawa takut dikira ngehujat padahal ...

Gambar 10. Hasil *Example Set* metode PSO+SVM

Hasil utama dari model optimasi PSO dan klasifikasi SVM adalah berupa hasil performansi klasifikasi. Adapun hasil

performansi dari model ini disampaikan sebagai berikut :

	true POSITIF	true NEGATIF	class precision
accuracy: 79.60% +/- 1.84% (micro average: 79.60%)			
pred. POSITIF	7	0	100.00%
pred. NEGATIF	102	381	79.31%
class recall	6.42%	100.00%	

Gambar 11. Hasil Performansi PSO+SVM

Dari hasil performansi model optimasi PSO dan klasifikasi SVM didapatkan data bahwa hasil akurasi pengujian SVM untuk analisis sentiment pada kolom komentar Instagram memperoleh hasil sebesar 79.60% +/- 1.84% (micro average: 79.60%). Untuk class precision prediksi positif adalah 100.00%, untuk class precision prediksi negatif sebesar 79.31% dan recall yang diperoleh dari true positif sebesar 6.42% sedangkan untuk true negatif sebesar 100.00%.

Berdasarkan hasil penelitian kali ini, maka didapatkan hasil keseluruhan dapat dilihat pada tabel 3 sebagai berikut :

Tabel 3 Hasil Penelitian

Metode	Class Recall (+)	Class Recall (-)	Akurasi
Naïve Bayes + PSO	0,00%	99,74%	78,00%
SVM + PSO	6,42%	100%	79,60%

4. Kesimpulan

Dari penelitian yang dilakukan pada penelitian kali ini, didapatkan beberapa kesimpulan. Kesimpulan pertama yang dapat diambil pada penelitian kali ini adalah metode *Particle Swarm Optimization* (PSO) merupakan salah satu metode optimasi yang dapat memberikan peningkatan hasil akurasi terhadap klasifikasi dengan jenis data dan beberapa metode klasifikasi yang menunjang. Pada penelitian kali ini, metode PSO memberikan hasil yang lebih baik jika

dikombinasikan dengan metode klasifikasi *Support Vector Machine* (SVM) jika dibandingkan dengan metode klasifikasi naïve bayes. Hal ini dibuktikan dengan hasil akurasi yang didapatkan oleh metode SVM yang didukung oleh metode optimasi PSO mencapai hasil akurasi sebesar 79,60% dengan dukungan *precision class* sebanyak 100%. Sedangkan untuk metode naïve bayes dengan dukungan metode PSO hanya mencapai nilai akurasi sebesar 78,00%.

Referensi

- Ardianto, R., Rivanie, T., Alkhalifi, Y., Nugraha, F. S., & Gata, W. (2020). Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine. *Jurnal Ilmu Komputer dan Informasi*, Vol.13, No.2, 109-121.
- Deolika, A., Kusriani, K., & Luthfi, E. T. (2019). Analisis Pemobotan Kata Pada Klasifikasi Text Mining. *JURTI (Jurnal Teknologi Informasi)*, 179-184.
- Destitus, C., Wella, & Suryasari. (2020). Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia. *ULTIMA InfoSys Vol.XI, No.2*, 107-111.
- Ernamia, E. A., & Herliana, A. (2022). Analisis Sentimen Kuliah Daring Dengan Algoritma Naïve Bayes, K-Nn Dan Decision Tree. *Jurnal Responsif*, 70-80.

- Fahlevi, F., & Sutriyanto, E. (2023, Februari 1). 1.895 Remaja Alami Perundungan Secara Siber, Pelakunya 1.182 Siswa. Diambil kembali dari [tribunnews.com: https://www.tribunnews.com/nasiona/2023/02/01/1895-remaja-alami-perundungan-secara-siber-pelakunya-1182-siswa](https://www.tribunnews.com/nasiona/2023/02/01/1895-remaja-alami-perundungan-secara-siber-pelakunya-1182-siswa)
- Fanny, F., Muliono, Y., & Tanzil, F. (2018). A Comparison of Text Classification Methods k-NN, Naïve Bayes, and Support Vector Machine for News Classification. *Jurnal Informatika : Jurnal Pengembangan IT (JPIT)*, Vol.03, No.02, 157-160.
- Jumeilah, F. S. (2017). Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* Vol.1, No.1, 19-25.
- Maulana, F. A., & Ernawati, I. (2020). Analisa Sentimen Cyberbullying di Jejaring Sosial Twitter Dengan Algoritma Naive Bayes. *SENAMIKA (Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya)* (hal. 529-538). Jakarta: Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
- Muhammad, A. N., Bukhori, S., & Pandunata, P. (2019). Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes – Support Vector Machine (NBSVM) Classifier. *International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)* (hal. 199-205). Jember - Indonesia: IEEE.
- Mustajab, R. (2021, Januari 31). *Indonesia Miliki 97,17 Juta Pengguna Instagram hingga Akhir 2022*. Diambil kembali dari [dataindonesia.id: https://dataindonesia.id/internet/data/indonesia-miliki-9717-juta-pengguna-instagram-hingga-akhir-2022](https://dataindonesia.id/internet/data/indonesia-miliki-9717-juta-pengguna-instagram-hingga-akhir-2022)
- Pajri, D., Umidah, Y., & Padilah, T. N. (2020). K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia. *Jurnal Teknik Informatika dan Sistem Informasi (JuTISI)* Vol.6, No.2, 242-253.
- Pratama, K. R., & Nistanto, R. K. (2021, Maret 29). *Instagram, Media Sosial Pemicu "Cyberbullying" Tertinggi*. Diambil kembali dari [tekno.kompas.com: https://tekno.kompas.com/read/2021/03/29/07164137/instagram-media-sosial-pemicu-cyberbullying-tertinggi#google_vignette](https://tekno.kompas.com/read/2021/03/29/07164137/instagram-media-sosial-pemicu-cyberbullying-tertinggi#google_vignette)
- Priyanto, A., & Ma'arif, M. R. (2018). Implementasi Web Scraping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi Laman Web Tentang Hidroponik. *Indonesian Journal of Information Systems (IJIS)* Vol.1, No.1, 25-33.
- Rachmayanti, A., & Candrasari, Y. (2022). Perilaku Cybebullying di Instagram. *LINIMASA : Jurnal Ilmu Komunikasi* Vol.5, No.1, 1-12.
- Rizkiansyah, Herliana, A., Alamsyah, D. P., & Tjoe, T. F. (2022). Comparison of the K-Nearest Neighbor and Decision Tree algorithm to the Sentiment Analysis of Investment Applications Users in Indonesia. *Seventh International Conference on Informatics and Computing (ICIC)* (hal. 1-6). Denpasar, Bali: IEEE.
- Sugiyono. (2018). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.
- Tysara, L. (2023, Mei 30). *Jenis Penelitian Kualitatif Menurut Para Ahli, Pahami Karakteristiknya*. Diambil kembali dari [liputan6.com: https://www.liputan6.com/hot/read/5299910/jenis-penelitian-kualitatif-menurut-para-ahli-pahami-karakteristiknya?page=4](https://www.liputan6.com/hot/read/5299910/jenis-penelitian-kualitatif-menurut-para-ahli-pahami-karakteristiknya?page=4)
- Watraton, A. F., Puspita, A., & Moeis, D. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. *Journal of Applied Computer Science And Technology (JACOST)* Vol.1, No.1, 7-14.