JURNAL RESPONSIF, Vol. 5 No.2 Agustus 2023, pp. 434~443
E-ISSN**:** 2685-6964

PENERAPAN *PARTICLE SWARM OPTIMIZATION* PADA DETEKSI UJARAN KEBENCIAN DALAM *PLATFORM TWITTER*

Mirza Yogy Kurniawan[1], Fathul Hafidh[2], Muhammad Edya Rosadi[3], Rezky Izzatul Yazidah Anwar[4]

[1]Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin
*****e-mail korespondensi: mirza.yogy@gmail.com

[2]Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin
e-mail: hafidh@uniska-bjm.ac.id

[3]Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin
e-mail: edya.rosadi@gmail.com

[4]Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin
e-mail: rezky.izzatul@uniska-bjm.ac.id

Abstrak

Media sosial telah menjadi sarana komunikasi yang sangat umum dipakai. Media sosial daring kini telah luas digunakan untuk pemasaran, pariwisata, dan juga penyebaran berita. Tingginya angka pengguna media sosial juga meningkatkan kemungkinan terdapatnya ujaran kebencian pada media tersebut. Ujaran kebencian adalah suatu bentuk tindakan komunikasi dalam bentuk provokasi, hasutan, atau hinaan kepada suatu individu atau kelompok dalam hal suku, agama, ras, kewarganegaraan dan lainnya. Riset ini mengumpulkan data cuitan dari Twitter menggunakan *Application Programming Interface* yang sudah disediakan oleh Twitter. Cuitan tersebut disimpan dan ditandai secara manual mana yang termasuk ujaran kebencian, dan mana yang bukan. Data cuitan diklasifikasi menggunakan metode *Random Forest, k-Nearest Neighbour, Naïve Bayes dan Support Vector Machine*. Hasil klasifikasi dicatat dan dilanjutkan dengan penerapan Particle Swarm Optimiziation. Hasil penerapan PSO dibandingkan dan didapati bahwa nilai akurasi yang tertinggi didapat oleh metode SVM yang dioptimasi dengan PSO dengan nilai akurasi sebesar 75.51% Sedangkan metode *Naïve Bayes* yang memiliki nilai F1 Score terbesar yaitu 64.87%. Pengaruh penerapan PSO terbesar terdapat pada klasifikasi PSO yang meningkat nilai F1 Scorenya dengan selisih 9.2%.

Kata Kunci: SVM, PSO, *Naïve Bayes*, Optimasi, Ujaran Kebencian

Abstract

Social media has become a very common form of communication. Online social media is now widely used for marketing, tourism, and news. The high number of social media users also increases the likelihood of hate speech on these media. Hate speech is an act of communication in the form of provocation, incitement, or insult to an individual or group in terms of ethnicity, religion, race, nationality and others. This research collect tweet data from Twitter using the Application Programming Interface provided by Twitter. The tweets are stored and manually marked which ones are hate speech, and which ones are not. The tweet data is classified using Random Forest, k-Nearest Neighbor, Naïve Bayes and Support Vector Machine methods. The classification results are recorded and optimized with Particle Swarm Optimiziation. The results of the application of PSO were compared and found that the highest accuracy value was obtained by the SVM method optimized with PSO with an accuracy value of 75.51% while the Naïve Bayes method had the largest F1 Score value of 64.87%. The biggest effect of PSO application is on PSO classification which increases the F1 Score value by a difference of 9.2%.

Keywords: *SVM, PSO, Naïve Bayes, Optimization, Hate Speech*

Naskah diterima 20 Agustus 2023; direvisi 30 Agustus 2023; diterbitkan 31 Agustus 2023

1. Pendahuluan

Ujaran kebencian adalah suatu bentuk tindakan komunikasi dalam bentuk provokasi, hasutan, atau hinaan kepada suatu individu atau kelompok dalam hal suku, agama, ras, kewarganegaraan dan lainnya. Ujaran kebencian didefinisikan oleh (Djuric et al., 2015) sebagai ucapan kasar yang ditujukan kepada kelompok dengan karakteristik tertentu seperti etnis, agama, dan gender, yang mana ucapan ini dapat mengganggu, memberikan kesan negatif pada bisnis daring dan kenyamanan pengguna halaman web.

Media sosial telah menjadi sarana komunikasi yang sangat umum dipakai. Media sosial daring kini telah luas digunakan untuk pemasaran (Satyadewi et al., 2017), pariwisata (Rathore et al., 2017), dan juga penyebaran berita (Paramastri & Gumilar, 2019). Tingginya angka pengguna media sosial juga meningkatkan kemungkinan terdapatnya ujaran kebencian pada media tersebut. Pelaku ujaran kebencian bisa berasal dari masyarakat umum, pelajar, mahasiswa, hingga tokoh ternama (Riswani et al., 2019).

Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE) yang kemudian berubah menjadi Undang-Undang Nomor 19 Tahun 2016 telah mengatur penggunaan teknologi informatika dan komputer (TIK), termasuk didalamnya tentang larangan menyebarkan berita bohong dan atau ujaran kebencian terhadap suku, agama, ras, dan antargolongan. Permasalahan yang timbul adalah bagaimana menyediakan tindakan preventif terhadap terjadinya tindakan tersebut dengan melibatkan TIK.

Penggunaan media sosial Twitter sebagai objek penelitian analisis sentimen telah dilakukan oleh (Antinasari et al., 2017) menggunakan metode *Naïve Bayes* pada data opini film, sedangkan (Alita et al., 2019) berfokus menganalisis pengaruh *emoticon* dan *sarcasm* untuk meningkatkan akurasi deteksinya dengan membandingkan dua metode yaitu SVM dan *Naïve Bayes*. Penggunaan SVM untuk analisis sentimen juga juga bisa didapati pada (Arsi et al., 2021) dan (Nuris et al., 2021) yang keduanya menerapkan optimasi PSO untuk meningkat akurasi dari SVM itu sendiri, dan terbukti akurasinya meningkat

Deteksi ujaran kebencian dengan Bahasa Indonesia pada media Twitter telah dilakukan oleh (Alfina et al., 2018) sedangkan deteksi penggunaan bahasa kasar telah dilakukan oleh (Ibrohim & Budi, 2018), keduanya menghasilkan *dataset* dan studi awal menggunakan *machine learning*.

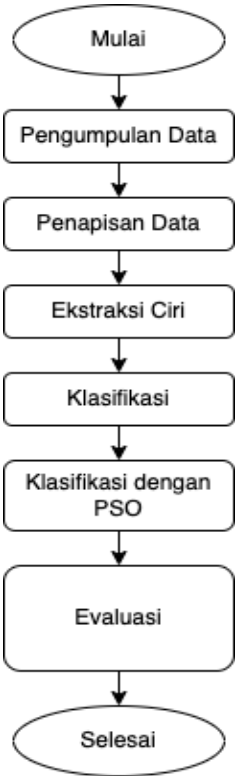
Penerapan metode SVM dan pembandingnya *Naïve Bayes* untuk data *hate speech* bisa didapati pada (Wenando, 2019), dan (Asogwa DC et al., 2022). Keduanya memiliki hasil yang berbeda karena ada perbedaan pada bagian ekstraksi cirinya. Metode lain yang digunakan untuk deteksi *hate speech* diantaranya ada (Taradhita & Putra, 2021) yang menggunakan *Convolutional Neural Network* dan (Kapil & Ekbal, 2020) yang menerapkan *Multi-task learning* berbasis *Deep Neural Network*

Riset ini mengumpulkan data cuitan dari Twitter menggunakan API yang sudah disediakan oleh Twitter. Cuitan tersebut disimpan beserta dengan biodata penulisnya yang berdasarkan dari kata pencarian jabatan kepala daerah, nama kepala daerah, kata kasar, dan hal yang terkait bencana. Data cuitan ditandai secara manual mana yang termasuk ujaran kebencian, dan mana yang bukan. Kemudian data yang sudah memiliki label dikonversi menjadi menjadi atribut untuk tiap data, dan diterapkan proses TF-IDF untuk ekstraksi cirinya. Hasil ekstraksi ciri diproses menggunakan metode *Naïve Bayes, Random Forest*, k-NN, dan *Support Vector Machine* untuk kemudian diukur tingkat akurasi, *precision*, dan *recall*-nya menggunakan *Cross Validation*. Berikutnya diterapkan PSO kepada setiap klasifikasi tersebut dan kembali diukur akurasi, *precision*, dan *recall*-nya untuk diamati peningkatan tiap nilai tersebut pada masing-masing metode klasifikasi.

2. Metode Penelitian

Penelitian ini melalui beberapa tahapan yang dilakukan agar dapat menghasilkan aplikasi yang mampu menyelesaikan permasalahan pada deteksi ujaran kebencian

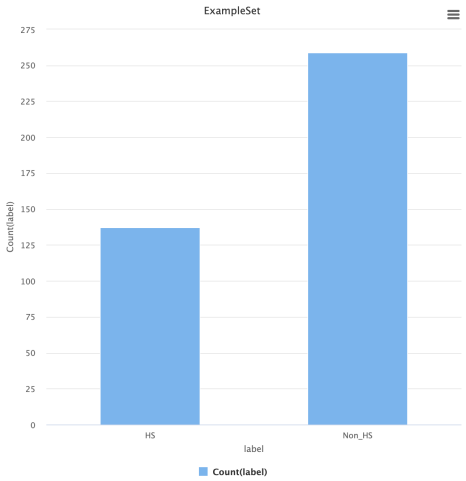
Tahapan yang telah dikerjakan digambarkan pada gambar 1 berikut:



Gambar 1. Metode Penelitian

Berdasarkan gambar 1 Tahapan dimulai dengan pengumpulan data pada aplikasi web yang mengakses data Twitter melalui API. Kemudian dilakukan penapisan data yang sama, atau berupa URL, dan data cuitan berupa teks diekstraksi cirinya untuk memudahkan proses klasifikasi. Tahapan berikutnya klasifikasi dioptimasi menggunakan petode PSO, dan pada tahap terakhir atau evaluasi, hasil antara klasifikasi tanpa PSO dengan klasifikasi berbasis PSO dibandingkan.

1. Pembangunan aplikasi web PHP yang menggunakan Twitter *Streaming Application Programming Interface* (API) untuk mengambil data cuitan. Data diambil dengan menggunakan kata kunci yang relevan dengan ujaran kebencian. Data diambil dalam rentang waktu 7 hari, pada jam-jam yang menurut (Spasojevic et al., 2015) merupakan jam paling padat yaitu jam 10.00-12.00 dan 19.00-21.00.
2. Data yang didapat disaring dari data yang terduplikasi, *retweet*, dan URL. Seluruh data teks juga diubah menjadi huruf kecil. Akhirnya didapat 396 baris data.
3. Proses pelabelan pada data yang sudah disaring secara manual dan menghasilkan 137 baris dilabeli *Hate Speech* (HS) sedangkan 259 sisanya bukan *Hate Speech* (Non_HS)



Gambar 2. Perbandingan jumlah dataset HS dengan Non_HS

4. Ekstraksi ciri yang digunakan pada penelitian ini adalah *Term Frequency-Inverse Document Frequency* (TF-IDF).
5. Metode klasifikasi yang dibandingkan adalah *Random Forest, Naïve Bayes*, k-NN, dan SVM. Masing-masing klasifikasi akan diukur akurasi, *precision, recall, dan F1 Score*
6. Masing-masing klasifikasi akan dioptimasi menggunakan PSO
7. Hasil klasifikasi saling dibandingkan akurasinya.

3. Hasil dan Pembahasan

Berikut hasil penelitian yang telah dilaksanakan

3.1. Pengumpulan Data

Tahap pengumpulan data menggunakan Twitter API pada tahun 2020 dengan kata kunci ‘jakarta banjir’, ‘jakarta macet’, ‘gubernur’, ‘anies’, ‘anies baswedan’, ‘ganjar’, ‘ganjar pranowo’, ‘ridwan kamil’, ‘bupati’, ‘walikota’, ‘presiden’. Hasil pengumpulan data ini didapati 22.320 baris data.

3.2. Penapisan Data

Sebanyak 22.320 baris data yang telah dihasilkan memiliki banyak data yang perlu dipilaj, sehingga proses dilanjutkan dengan pemilahan data yang terduplikasi, data yang merupakan retweet, dan data yang menggunakan, yang kemudian disaring dan dilabeli menghasilkan total 396 baris data.

Pada proses ini dilakukan juga text processing dimana setiap karakter “@” diganti dengan kata “USER_”, dilanjutkan dengan penghapusan karakter “?”, “!”,”.”, dan “,”. Terakhir kata negasi yang dituliskan dengan kata “ga”,”gak”,”engga”, “enggak”, “tak”, dan “tidak” diganti menjadi kata “NOT_”

Tabel 1. Sampel data

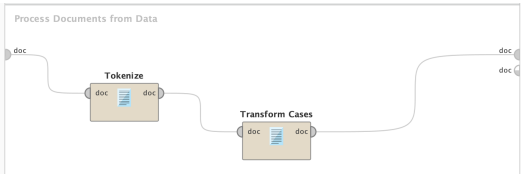
label	Text
HS	USER_mus_tanjung: Neh biar kalen2 bacin otak konslet paham kenapa gubernur indonesia USER_aniesbaswedan memilih monas sbg sirkuit formula E
HS	USER_Paltiwest: #WajahBaruJakarta setelah dipimpin Gubernur Terbodoh Silahkan tertawakan
HS	USER_FerdinandHaean2: Parah Ini gubernur mmg banyak omong sih Kinerja NOL
…	…
Non_HS	Sekalian Pak USER_aniesbaswedan kalo ditanya knapa Jakarta Banjir mulu jgn jawab tunggu air laut surut dong Skalian
Non_HS	USER_AriestaRiico: Jakarta Banjir lagi Bodo amat Sorry lagi sibuk ngurusin FormulaE Kok gtu Wan Sebenernya niat tidak sih jd Gabener
Non_HS	Inget Besok SENIN Malah jakarta Ujan mulu ni dari malam kalau sampe Pagi masih Hujan besok pasti pada Kena Macet karena Banjir

3.3. Ekstraksi Ciri

Metode ekstraksi ciri yang digunakan pada penelitian ini adalah TF-IDF (*Term Frequency-Inverse Document Frequency*) yang biasa digunakan dalam pemrosesan bahasa alami dan pengambilan informasi. Metode ini dapat membantu menemukan fitur atau ciri yang penting dari sebuah dokumen yang kemudian membantu proses klasifikasi data teks. TF-IDF juga secara umum diketahui memiliki performa yang baik ketika diterapkan untuk klasifikasi menggunakan SVM (Cahyani & Patasik, 2021).

Pada data diterapkan juga metode *Tokenize* untuk memberi token terkait kata yang ada pada data tersebut, selain itu untuk menyeragamkan data, seluruh kata juga diterapkan *Tranform Cases* yang berfungsi untuk mengubah semua karakter menjadi huruf kecil, karena pada pengenalan kata huruf besar dan huruf kecil dianggap sebagai kata yang berbeda.

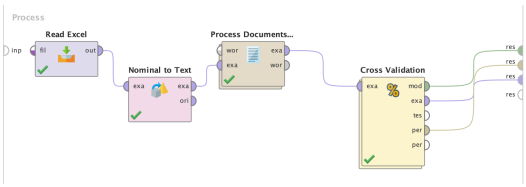
Berikut adalah gambar implementasi dari ekstraksi ciri yang telah dijelaskan.



Gambar 3. Tahapan Ekstraksi Ciri

3.4. Klasifikasi Tanpa PSO

Metode klasifikasi yang digunakan pada penelitian ini adalah *Naïve Bayes*, *Support Vector Machine, k-Nearest Neighbour, dan Random Forest*. Pada tiap klasifikasi diukur tingkat akurasi, presisi, dan recall-nya.

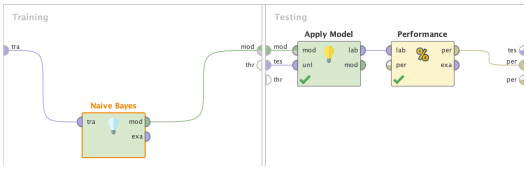


Gambar 4. Penerapan *Cross Validation*

Pada gambar 4 dapat dilihat Data yang dihasilkan dari penapisan data dikonversi kedalam bentuk tabel Excel kemudian diimpor ke dalam blok *Read Excel*, data juga diproses agar didapat nilai TF-IDF nya.

1. Klasifikasi Naïve Bayes

Pada gambar 5 yang merupakan gambar detail dari blok *Cross Validation* berisikan blok metode klasifikasi *Naïve Bayes* yang diukur performanya dengan menggunakan *10-fold cross validation*. Pengukuran ini akan menghasilkan nilai Akurasi, *Precision*, dan *Recall*.



Gambar 5. Penerapan Klasifikasi *Naïve Bayes*

Tabel 2 menyajikan hasil validasi klasifikasi *Naïve Bayes* terhadap data *hate speech*

Tabel 2. Hasil Validasi Naïve Bayes

N = 396	*True HS*	*True Non_HS*	*Class Precision*
Pred. HS	85	68	55.56%
Pred. Non_HS	52	191	78.60%
Class recall	62.04%	73.75%	

Berdasar tabel 2 yang biasa disebut dengan *confusion matrix* dapat disusun nilai akurasi, presisi, *recall,* dan F1 *score* dengan cara identifikasi terlebih dahulu nilai *True Positive, True Negative, False Positive,* dan *False Negative*.

True Positive (TP) merupakan jumlah data yang berhasil diklasifikasikan oleh metode sebagai HS dan sesuai dengan labelnya yaitu HS, pada tabel 2 didapati bahwa nilai TP adalah 85.

True Negative (TN) adalah jumlah data yang berhasil diklasifikasikan metode sebagai Non_HS dan sesuai dengan labelnya bahwa data tersebut Non_HS. Pada tabel 2 nilai TN adalah 191.

False Positive (FP) merupakan jumlah data yang diklasifikasikan oleh metode sebagai HS, namun kenyataannya label pada data tersebut adalah Non_HS. Pada tabel 2 didapati nilai FP adalah 68.

False Negative (FN) adalah jumlah data yang diklasifikasikan oleh metode sebaga Non_HS tapi pada kenyataannya memiliki label HS. Pada tabel 2 nilai FN didapati sebesar 52

Akurasi = (TP + TN / N) * 100%
= (85+191 / 396) * 100%
= 69.7%

Precision = (TP/(TP+FP)) * 100%
= 85/(85+68) * 100%
= 55.56%

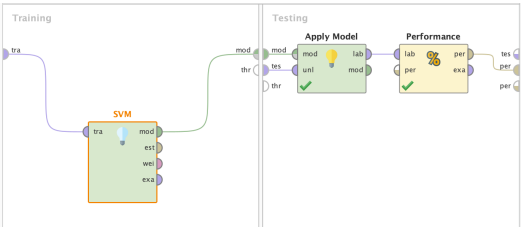
Recall = (TP/(TP+FN)) * 100%
= 85/(85+52) * 100%
= 62.04%

F1 *Score* = 2*(*Recall*Precision*) / (*Recall*+*Precision*))
= 2*(62.04*55.56)/(62.04+55.56))
= 58.62

Nilai akurasi mewakili bagaimana performa klasifikasi terhadap dataset yang digunakan secara keserluruhan, sedangkan presisi mengukurnya terhadap label positif dalam hal ini label positif yang digunakan adalah HS. Sedangkan *recall* menggambar kan performa klasifikasi terhadap label negatif yaitu Non_HS. F1 *Score* merupakan rataan harmonik dari presisi dan *recall*. Nilai ini dianggap lebih baik dalam mengukur performa pada dataset yang tidak seimbang labelnya.

2. Klasifikasi SVM

Pada Gambar 6 menunjukkan implementasi SVM pada data *hate speech* yang divalidasi menggunakan *Cross Validation*



Gambar 6. Penerapan Klasifikasi SVM

Proses penggantian klasifikasi pada rapidminer cukup dilakukan dengan mengganti blok model yang tadinya Naïve Bayes dengan blok SVM, kemudian disambungkan Kembali

Tabel 3. Hasil Validasi SVM

N = 396	*True HS*	*True Non_HS*	*Class Precision*
Pred. HS	32	3	91.43%
Pred. Non_HS	105	256	70.91%
Class recall	23.36%	98.84%	

Berdasarkan tabel 3 disusun pengukuran performa klasifikasi SVM sebagai berikut.

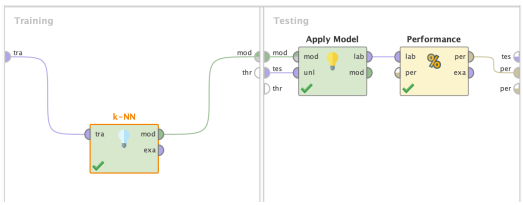
Akurasi = (TP + TN / N) * 100%
= 32+256 / 396 * 100%
= 72.73%
Precision = (TP/(TP+FP)) * 100%
= (32/(32+3)) * 100%
= 91.43%

Recall = (TP/(TP+FN)) * 100%
= (32/(32+105)) * 100%
= 23.36%
F1 *Score* = (2*(*Recall*Precision*) / (*Recall*+*Precision*)) * 100%
= 2*(23.36*91.43)/(23.36+91.43)
= 37.21

Pada performa SVM ini dapat diamati akurasinya lebih tinggi daripada Naïve Bayes, begitu juga dengan presisinya, namun ada masalah pada nilai *recall*-nya dimana ini terjadi karena banyaknya data HS yang diklasifikasi sebagai Non_HS

3. Klasifikasi k-NN

Pada Gambar 7 menunjukkan implementasi k-NN pada data *hate speech* yang divalidasi menggunakan *Cross Validation*



Gambar 7. Penerapan Klasifikasi k-NN

Tampilan pada gambar 7 menunjukkan hasil penggantian blok SVM dengan blok model k-NN.

Tabel 4. Hasil Validasi k-NN

N = 396	*True HS*	*True Non_HS*	*Class Precision*
Pred. HS	74	39	65.49%
Pred. Non_HS	63	220	77.74%
Class recall	54.01%	84.94%	

Berdasarkan hasil klasifikasi yang dituangkan dalam tabel 4 maka diukur performa k-NN dalam klasifikasi ini.

Akurasi = (TP + TN / N) * 100%
= (74+220 / 396) * 100%
= 74.24%

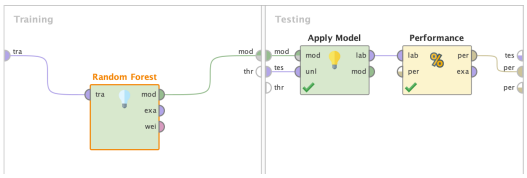
Precision = (TP/(TP+FP)) * 100%
= 74/(74+39)
= 65.49%

Recall = (TP/(TP+FN)) * 100%
= (74/(74+63)) * 100%
= 54.01%

F1 *Score* = 2*(*Recall***Precision*) / (*Recall*+*Precision*)
= 2*(54.014*65.49)/(54.01+65.49)
= 59.2%

4. Klasifikasi *Random Forest*

Random Forest digunakan pada eksperimen ini dengan cara mengganti blok model sebelumnya dengan blok *Random Forest*



Gambar 8. Penerapan Klasifikasi RF

Tangkapan layar pada gambar 8 menunjukkan penerapan *Random Forest* pada aplikasi Rapidminer

Tabel 5. Hasil Validasi *Random Forest*

N = 396	*True HS*	*True Non_HS*	*Class Precision*
Pred. HS	0	0	0%
Pred. Non_HS	137	259	65.40%
Class recall	0%	100%	

Berdasarkan penyajian hasil pada tabel 5 didapat bahwa tidak ada satupun data yang berhasil diklasifikasi sebagai HS, semua data dianggap sebagai Non_HS. Berikut perhitungan performanya

Akurasi = (TP + TN / N) * 100%
= (0+259 / 396) * 100%
= 65.4%

Precision = (TP/(TP+FP)) * 100%
= (0/(0+0)) * 100%
= Tidak Valid

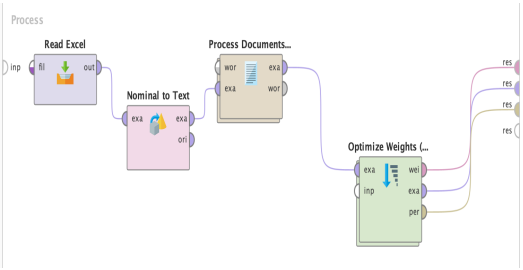
Recall = (TP/(TP+FN)) * 100%
= (0/(0+137)) * 100%
= 0

F1 *Score* = 2*(*Recall***Precision*) / (*Recall*+*Precision*)
= Tidak Valid

Meskipun memiliki nilai akurasi 65.4% namun memiliki nilai presisi yang tidak valid karena pembagian dengan nilai 0, begitu juga dengan F1 *score*-nya karena F1 *score* berkaitan langsung dengan nilai presisi.

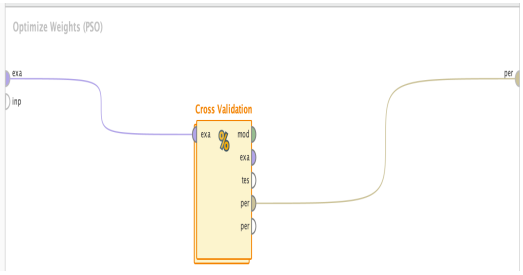
3.4. Klasifikasi dengan PSO

Klasifikasi dengan PSO dilakukan dengan menambahkan blok PSO dan memindahkan *Cross Validation* ke dalam blok tersebut sesuai yang ditunjukkan pada gambar 8



Gambar 9. Penerapan PSO

Blok validasi sebagaimana yang sudah ada pada gambar 4, dipindahkan ke dalam blok PSO dan dapat dilihat pada gambar 10



Gambar 10. *Cross Validation* dalam PSO

1. Klasifikasi Naïve Bayes + PSO

Tabel 6 menyajikan hasil klasifikasi menggunakan *Naïve Bayes* yang sudah dioptimasi menggunakan PSO

Tabel 6. Hasil Validasi *Naïve Bayes* + PSO

	True HS	*True Non_HS*
Pred. HS	96	63
Pred. Non_HS	41	196

Berdasarkan tabel 6, disusun perhitungan performa klasifikasi Naïve Bayes yang dioptimasi dengan PSO

Akurasi = (TP + TN / N) * 100%
= (96+196 / 396) * 100%
= 73.74*

Precision = (TP/(TP+FP)) * 100%
= (96/(96+63)) * 100%
= 60.38%

Recall = (TP/(TP+FN)) * 100%
= (96/(96+41)) * 100%
= 70.07%

F1 *Score* = (2*(*Recall***Precision*) / (Recall+Precision)
= 2*(70.07*60.38)/(70.07+60.38)
= 64.87

Berdasarkan pengukuran tersebut dapat diamati bahwa pada penerapan PSO pada metode *Naïve Bayes* dapat meningkatkan performa klasifikasi tersebut, dapat diamati dari meningkatnya seluruh nilai performa tersebut terutama F1 *score* yang tadinya 58.62 meningkat menjadi 64.87 dengan peningkatan sebesar 6.25

2. Klasifikasi SVM + PSO

Tabel 7 menyajikan hasil klasifikasi menggunakan *Suport Vector Machine* yang sudah dioptimasi menggunakan PSO

Tabel 7. Hasil Validasi SVM + PSO

	True HS	*True Non_HS*
Pred. HS	42	2
Pred. Non_HS	95	257

Akurasi = (TP + TN / N) * 100%
= (42+257 / 396) * 100%
= 75.51%

Precision = (TP/(TP+FP) * 100%
= (42/(42+2)) * 100 %
= 95.45%

Recall = (TP/(TP+FN)) * 100%
= (42/(42+95)) * 100%
= 30.66%

F1 *Score* = 2*(*Recall***Precision*) / (*Recall*+*Precision*)
= 2*(30.66*95.45)/(30.66+95.45)
= 46.41

Pengukuran performa SVM yang dioptimasi dengan PSO menunjukkan peningkatan untuk seluruh nilainya, terutama F1 *Score* yang memiliki selisih 9.2 dibandingkan dengan SVM tanpa PSO. Meskipun demikian, nilai ini memang didapati sangat rendah.

3. Klasifikasi k-NN + PSO

Tabel 8 menyajikan hasil klasifikasi menggunakan *k-Nearest Neighbour* yang sudah dioptimasi menggunakan PSO

Tabel 8. Hasil Validasi k-NN + PSO

	True HS	*True Non_HS*
Pred. HS	67	31
Pred. Non_HS	70	228

Akurasi = (TP + TN / N) * 100%
= (67+228/396) * 100%
= 74.49%

Precision = (TP/(TP+FP)) * 100%
= (67/(67+31)) * 100%
= 68.37%

Recall = (TP/(TP+FN)) * 100%
= (67/(67+70))) * 100%
= 48.91

F1 *Score* = 2*(*Recall*Precision*) / (*Recall*+*Precision*)
= 2*(48.91*48.91)/(0.49+68.37)
= 57.03

Performa k-NN yang dioptimasi dengan PSO diukur dan didapati bahwa meskipun akurasi dan presisinya meningkat, namun terdapat penurunan *recall* sehingga berdampak pada menurunnya F1 *score.* Hal ini disebabkan oleh pembobotan PSO yang justru menurunkan performa k-NN untuk mengklasifikasi data

4. Klasifikasi *Random Forest* + PSO

Hasil klasifikasi Random Forest dengan optimasi PSO memiliki hasil yang sama dengan Tabel 5, sehingga nilai evaluasinya menjadi tidak bisa digunakan

3.5. Evaluasi

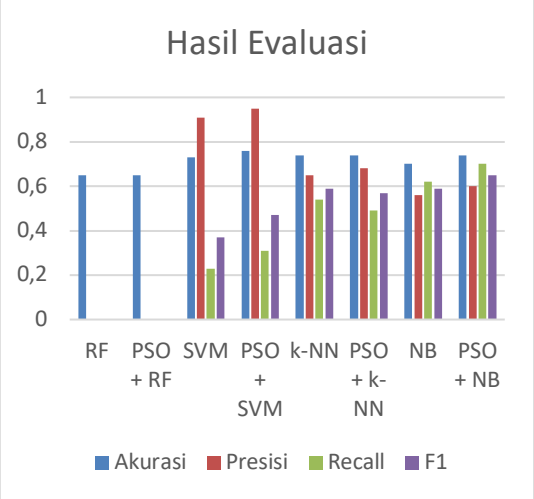
Hasil klasifikasi dari tiap metode dicatat dan dibandingkan hasilnya dan disajikan pada Tabel 9.

Tabel 9. Evaluasi Hasil Klasifikasi

Nilai	Akurasi	Presisi	*Recall*	F1
Random Forest	65.4	Tidak Valid	0	Tidak Valid
PSO *Random Forest*	65.4	Tidak Valid	0	Tidak Valid
SVM	72.73	91.43	23.36	37.21
PSO SVM	75.51	95.45	30.66	46.41
k-NN	74.24	65.49	54.01	59.2
PSO k-NN	74.49	68.37	48.91	57.03
Naïve Bayes	69.7	55.56	62.04	58.62
PSO *Naïve Bayes*	73.74	60.38	70.07	64.87

Performa hasil klasifikasi tiap metode beserta performa klasifikasi yang dioptimasi dengan PSO yang tampil pada tabel 9 menunjukkan bahwa PSO dapat meningkatkan performa. Meskipun untuk kasus k-NN, PSO justru menurunkan performanya. Peningkatan dapat dilihat pada metode *Naïve Bayes* yang memiliki selisih F1 sebesar 6.25 dan SVM yang memiliki selisih sebesar 9.2.

Berikut gambar yangmenyajikan hasil evaluasi dalam bentuk diagram batang



Gambar 11. Grafik Hasil Evaluasi

Grafik pada gambar 10 menunjukkan perbandingan performa pada tiap metode klasifikasi, baik sebelum dan juga sesudah penerapan optimasi PSO.

4. Kesimpulan

Metode optimasi PSO yang diterapkan pada klasifikasi data ujaran kebencian dapat meningkatkan performa pada metode *Naïve Bayes* dan SVM, sedangkan pada k-NN justru menurunkan performanya.

Berdasarkan tabel 9 dapat diamati bahwa akurasi yang tertinggi didapat oleh metode PSO SVM dengan akurasi sebesar 75.51% namun karena data yang digunakan bersifat *imbalance* maka yang menjadi perhatian adalah F1 *Score* dimana nilai tertingginya adalah Metode PSO NB dengan nilai sebesar 64.87%. Hasil ini menunjukkan Naïve Bayes merupakan metode yang paling cocok untuk dataset pada penelitian ini. Sedangkan terkait dengan dampak PSO pada performa klasifikasi dapat dilihat bahwa pada klasifikasi SVM metode PSO dapat meningkatkan performa yang paling besar yaitu didapati selisih F1 *score* sebesar 9.2%.

Referensi

Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2018). Hate speech detection in the Indonesian language: A dataset and preliminary study. *2017 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2017*, *2018-Janua*(October), 233–237. https://doi.org/10.1109/ICACSIS.2017.8355039

Alita, D., Priyanta, S., & Rokhman, N. (2019). Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter. *Journal of Information Systems Engineering and Business Intelligence*, *5*(2), 100. https://doi.org/10.20473/jisebi.5.2.100-109

Antinasari, P., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(12), 1718–1724. http://j-ptiik.ub.ac.id

Arsi, P., Wahyudi, R., & Waluyo, R. (2021). Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(2), 231–237. https://doi.org/10.29207/resti.v5i2.2698

Asogwa DC, Chukwuneke CI, Ngene CC, & Anigbogu GN. (2022). Hate Speech Classification Using SVM and Naive BAYES. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA)*, *9*(1), 27–34. https://doi.org/10.9790/0050-09012734

Cahyani, D. E., & Patasik, I. (2021). Performance comparison of tf-idf and word2vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, *10*(5), 2780–2788. https://doi.org/10.11591/eei.v10i5.3157

Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V., & BhamidipatiNarayan. (2015). Hate Speech Detection with Comment Embeddings. *Proceedings of the 24th International Conference on World Wide Web*.

Ibrohim, M. O., & Budi, I. (2018). A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media. *Procedia Computer Science*, *135*, 222–229. https://doi.org/10.1016/j.procs.2018.08.169

Kapil, P., & Ekbal, A. (2020). A deep neural network based multi-task learning approach to hate speech detection. *Knowledge-Based Systems*, *210*, 106458. https://doi.org/10.1016/j.knosys.2020.106458

Nuris, N., Rini Yulia, E., Solecha, K., Bina, U., Infromatika, S., & Mandiri, U. N. (2021). Implementasi Particle Swarm Optimization (PSO) Pada Analysis Sentiment Review Aplikasi Halodoc Menggunakan Algoritma Naïve Bayes. *Jurnal Teknologi Informasi*, *7*. http://ejournal.urindo.ac.id/index.php/TI

Paramastri, N. A., & Gumilar, G. (2019). Penggunaan Twitter Sebagai Medium Distribusi Berita dan News Gathering Oleh Tirto.Id. *Jurnal Kajian Jurnalisme*, *3*(1), 18. https://doi.org/10.24198/jkj.v3i1.22450

Rathore, A. K., Joshi, U. C., & Ilavarasan, P. V. (2017). Social Media Usage for Tourism: A Case of Rajasthan Tourism. *Procedia Computer Science*, *122*, 751–758. https://doi.org/10.1016/j.procs.2017.11.433

Riswani, Khaidir, E., Suhertina, & Zaliana. (2019). Sikap Siswa tehadap Hate Speech dan Layanan Bimbingan Konseling di Sekolah Pada Era Revolusi 4.0. *Konvensi Nasional XXI Asosiasi Bimbingan Dan Konseling Indonesia*, *April*, 213–206.

Satyadewi, A. J., Hafiar, H., & Nugraha, A. R. (2017). Pemilihan Akun Media Sosial INSTAGRAM oleh HOLIDAY INN Bandung. *Jurnal The Messenger*, *9*(2), 153. https://doi.org/10.26623/themessenger.v9i2.459

Spasojevic, N., Li, Z., Rao, A., & Bhattacharyya, P. (2015). When-to-post on social networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *2015-Augus*, 2127–2136. https://doi.org/10.1145/2783258.2788584

Taradhita, D. A. N., & Putra, I. K. G. D. (2021). Hate speech classification in Indonesian language tweets by using convolutional neural network. *Journal of ICT Research and Applications*, *14*(3), 225–239. https://doi.org/10.5614/itbj.ict.res.appl.2021.14.3.2

Wenando, F. A. (2019). Detection of Hate Speech in Indonesian Language on Twitter Using Machine Learning Algorithm. *PROCEEDING CelSciTech-UMRI 2019*, *4*, 6–8.