

# Penerapan *Binary Particle Swarm Optimization Support Vector Machine* untuk Klasifikasi Komentar *Cyberbullying* di Instagram \*

Dewi Fortuna<sup>1</sup>, Itasia Dina Sulvianti<sup>2‡</sup>, Gerry Alfa Dito<sup>3</sup>

<sup>123</sup>Department of Statistics, IPB University, Indonesia  
<sup>‡</sup>corresponding author: [itiasiasu@apps.ipb.ac.id](mailto:itiasiasu@apps.ipb.ac.id)

Copyright © 2022 Dewi Fortuna, Itasia Dina Sulvianti, and Gerry Alfa Dito. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

Freedom of speech on social media is sometimes inappropriate with the ethics of communicating and has led to cyberbullying. Instagram is the most commonly used social media in cyberbullying. Cyberbullying needs to be minimized because it has many adverse effects. One way that can be done is by identifying cyberbullying comments so those comments can be deleted automatically. The method used in this study is text classification using Support Vector Machine (SVM) algorithm with the application of Binary Particle Swarm Optimization (BPSO) optimization method as features selection. The study aims to build a cyberbullying comments classification model and compare the classification model performance with and without the application of features selection. The experimental results showed that modeling with SVM produces a reasonably accurate classification performance over 72% for all classification performance on each C. The application of BPSO for features selection can improve classification performance by increasing accuracy and specificity. However, the model without features selection on  $C = 0,1$  is chosen in this study case because it has the highest sensitivity with good accuracy and specificity that can detect cyberbullying comments more accurately.

**Keywords:** binary particle swarm optimization, cyberbullying, feature selection, support vector machine, text mining.

## 1. Pendahuluan

Media sosial saat ini menjadi kebutuhan bagi setiap orang. Menurut laporan Hootsuite (*We are Social*), pengguna media sosial di Indonesia pada Januari 2020 mencapai 160 juta pengguna atau setara dengan 59% dari total penduduk Indonesia. Salah satu manfaat media sosial yaitu sebagai media berdiskusi dan menyampaikan pendapat. Akan tetapi, adakalanya pendapat yang diberikan tidak diikuti dengan penerapan etika dalam berkomunikasi sehingga mengarah kepada tindakan *cyberbullying*.

Cyberbullying merupakan perilaku seseorang atau sekelompok orang dalam

---

\* Received: Sep 2021; Reviewed: Oct 2021; Published: Jan 2022

melakukan tindakan yang dapat menyakiti orang lain melalui komputer, telepon seluler, dan alat elektronik lainnya secara sengaja dan berulang kali (Rifauddin 2016). Menurut hasil survei dari Ditch The Label (2018), Instagram merupakan media sosial yang paling umum digunakan dalam melakukan *cyberbullying*. Tindakan *bullying* yang dilakukan yaitu memberi komentar negatif pada postingan tertentu, pesan pribadi tak bersahabat, atau menyebarkan postingan seseorang dengan mengolok-olok (Ditch The Label 2018).

Tindakan *cyberbullying* pada Instagram perlu diminimalisasi karena menimbulkan banyak dampak negatif. Salah satu cara yang dapat dilakukan yaitu dengan mengidentifikasi suatu komentar apakah termasuk *cyberbullying* atau tidak. Hal ini dapat dilakukan dengan menerapkan analisis *text mining* dengan metode klasifikasi teks. Klasifikasi komentar dilakukan agar komentar *cyberbullying* dapat terdeteksi secara otomatis dengan bantuan *machine learning* sehingga komentar yang termasuk *cyberbullying* dapat terhapus secara otomatis. Metode klasifikasi yang akan digunakan pada penelitian ini yaitu metode *Support Vector Machine* (SVM).

Algoritme SVM merupakan salah satu metode yang umum digunakan dalam klasifikasi teks. Algoritme ini mempunyai akurasi yang tinggi dan lebih tidak rentan terhadap *overfitting* dibandingkan metode lainnya (Han et al. 2012). Penelitian yang dilakukan oleh Chrismanto dan Lukito (2017) menunjukkan SVM menghasilkan tingkat akurasi 78,49% yang lebih baik dari metode Naïve Bayes (77,25%) pada identifikasi komentar spam di Instagram. Penelitian lainnya juga dilakukan oleh Luqyana et al. (2018) mengenai analisis sentimen *cyberbullying* di Instagram menggunakan metode SVM yang menghasilkan akurasi sebesar 90%.

Data teks terdiri dari banyak fitur/kata, padahal tidak semua fitur sesuai dengan hasil klasifikasi. Untuk itu, diperlukan seleksi fitur untuk memilih fitur penting yang relevan terhadap data (Maulida et al. 2016). Seleksi fitur yang digunakan pada penelitian ini yaitu *Binary Particle Swarm Optimization* (BPSO). Metode BPSO dipilih karena efektif digunakan pada data berdimensi tinggi (Widiyanto 2018), dalam hal ini terdiri dari banyak fitur/kata. Metode ini juga mudah digunakan, serta cepat konvergen dengan sedikit parameter *tuning* (Kurniawan M 2017). Berdasarkan penelitian yang dilakukan oleh Pajri et al. (2020), penerapan seleksi fitur BPSO dapat meningkatkan akurasi metode KNN dari 88,11% menjadi 97,9%. Penerapan BPSO untuk seleksi fitur dan SVM sebagai algoritme klasifikasi diharapkan dapat memberikan hasil maksimal dalam pengklasifikasian *cyberbullying* di Instagram. Tujuan dari penelitian ini yaitu untuk menyusun model klasifikasi komentar *cyberbullying* serta membandingkan performa kinerja model klasifikasi tanpa seleksi fitur dan dengan seleksi fitur.

## 2. Metodologi

### 2.1 Data

Data yang digunakan dalam penelitian ini merupakan data hasil *scraping* komentar pada postingan beberapa akun Instagram yang pernah menjadi korban *cyberbullying* yaitu @rahmawatikekeyiputricantikka23, @cimoyluv, @viavallen, @riaricis1795, dan @aurelie.hermansyah. Data yang diambil dari setiap akun merupakan komentar dari beberapa postingan saja yaitu postingan pada saat akun tersebut mengalami *cyberbullying*. Banyaknya postingan yang diambil dari setiap akun masing-masing dua postingan, kecuali pada akun @viavallen yang hanya diambil satu postingan. Hal ini karena telah dihapusnya beberapa postingan pada akun tersebut pada masa

pemberitaan *cyberbullying*. Proses pengambilan data komentar Instagram dilakukan melalui proses *scraping* dengan bantuan *tools Phantombuster.com*. Berdasarkan hasil *scraping*, dihasilkan data sebanyak 7866 komentar dari 9 postingan.

## 2.2 Metode Penelitian

Proses analisis data dilakukan dengan menggunakan *Software R 4.0.2* dan *python 3.7.10*. Prosedur analisis data yang dilakukan yaitu sebagai berikut:

### 1. Pemilahan Data

Pemilahan data diawali dengan menghapus data duplikat yang disebabkan oleh kesalahan pada proses *scraping*. Setelah itu, dilakukan pemilahan data komentar sesuai konteks penelitian. Komentar yang tidak sesuai dengan konteks penelitian yaitu komentar yang berisi promosi/jualan serta komentar yang tidak mempunyai arti yang jelas. Komentar pada data terdiri dari dua jenis yaitu komentar secara langsung terhadap isi postingan dan komentar yang berupa jawaban terhadap komentar orang lain. Komentar yang digunakan pada penelitian ini yaitu komentar terhadap postingan akun, sehingga komentar yang hanya menjawab komentar orang lain tidak diikutsertakan dalam penelitian. Karena pada saat proses *scraping* emotikon tidak terbaca dengan jelas, maka komentar yang hanya berisi emotikon saja dan tidak mengandung huruf a-z tidak akan diikutsertakan dalam penelitian. Proses pada tahap pemilahan data menyisakan sebanyak 4.369 data komentar.

### 2. Pelabelan Data

Proses pelabelan data dilakukan secara manual oleh seorang mahasiswa Departemen Sains Komunikasi dan Pengembangan Masyarakat (SKPM) IPB University. Setiap komentar Instagram akan diberi satu label atau kelas. Komentar akan diberi label "1" jika komentar tersebut mengandung kalimat yang mengarah kepada *cyberbullying* dan diberi label "0" apabila komentar tidak mengandung unsur *cyberbullying*. Untuk keperluan eksplorasi data, komentar yang termasuk *cyberbullying* diberi label jenis *bullying* apakah termasuk *Called Name (cn)*, *Threatened Physical Harm (tp)*, atau *Opinion Slammed (os)*. Setelah diberi label, dihasilkan 539 data yang termasuk *cyberbullying* berlabel "1" dan 3830 data yang tidak termasuk *cyberbullying* berlabel "0".

### 3. Text Pre-processing

Proses ini terdiri dari beberapa tahapan yaitu *case folding*, *filtering*, normalisasi, *stemming*, dan penghapusan *stopwords*. Normalisasi dilakukan menggunakan kamus dari *colloquial-indonesian-lexicon* yang berisi pasangan kata formal dan informal dalam Bahasa Indonesia. Proses *stemming* dilakukan menggunakan *Library Python Sastrawi*. Penghapusan *stopwords* Bahasa Indonesia dilakukan menggunakan *Library Python Sastrawi* dengan beberapa modifikasi sedangkan untuk *stopwords* dalam Bahasa Inggris berasal dari *Natural Language Toolkit (NLTK)*.

### 4. Eksplorasi Data

Setiap data komentar dikelompokkan berdasarkan nama akun untuk melihat distribusi label pada setiap akun. Selain itu, dilakukan juga eksplorasi jenis *bullying* dan eksplorasi pada banyaknya teks komentar. Data komentar yang telah melalui *text pre-processing* masuk ke tahap pembentukan awan kata untuk mengetahui kata-kata yang berkaitan dengan komentar *cyberbullying* dan bukan *cyberbullying*.

### 5. Pembobotan TF-IDF

Data yang sudah melalui *text pre-processing* akan dibuat menjadi vektor menggunakan perhitungan TF-IDF. Pembobotan dilakukan pada masing-masing data latih dan data uji pada tahap *cross validation*. Banyaknya fitur yang diperoleh pada

tahap ini berbeda pada setiap *fold*. Pada tahap ini rata-rata fitur yang diperoleh sebanyak 2333 fitur atau kata. Fitur-fitur ini yang akan digunakan dalam pemodelan.

## 6. Pemodelan Klasifikasi

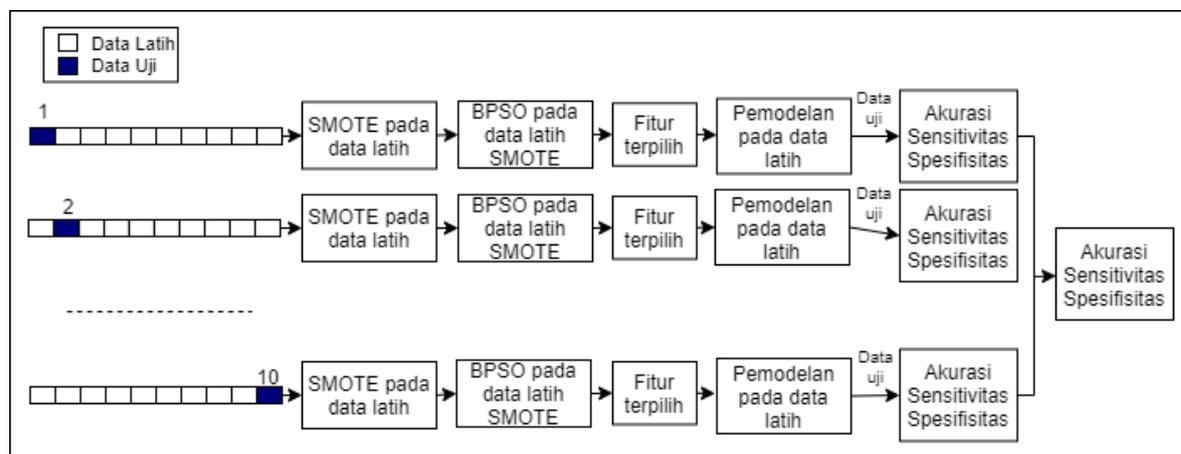
Pemodelan klasifikasi dilakukan dengan dua kondisi yaitu tanpa seleksi fitur dan dengan seleksi fitur BPSO.

### a. Pemodelan tanpa seleksi fitur

Pemodelan dilakukan menggunakan metode SVM dengan *10-fold cross-validation*. Menurut Joachims (1997), sebagian besar masalah klasifikasi teks adalah *linearly separable*. Maka dari itu, pada penelitian ini digunakan kernel linear dengan parameter C. Beberapa nilai parameter C yang akan diujikan berdasarkan penelitian dari Laia dan Setiawan (2020) yaitu 0,1; 1, dan 10. Karena data pada penelitian ini tidak seimbang, maka dilakukan penanganan ketidakseimbangan data menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Prosedur pemodelan tanpa seleksi fitur BPSO terdapat pada Gambar 1 hanya saja tanpa melibatkan seleksi fitur BPSO pada data latih SMOTE.

### b. Pemodelan dengan seleksi fitur BPSO

Fungsi *fitness* yang digunakan pada penelitian ini yaitu nilai *balanced accuracy*. Langkah-langkah pemilihan fitur dan pemodelan ditampilkan pada Gambar 1. Nilai parameter BPSO yang digunakan ditentukan dengan melihat penelitian sebelumnya dengan beberapa penyesuaian. Yan et al. (2016) menyatakan bahwa bobot inersia ( $w$ ) yang umum digunakan berkisar dari 0,4 sampai 0,9. Mafarja et al. (2018) pada penelitiannya menggunakan 10 partikel dan 100 iterasi menghasilkan hasil yang baik.



Gambar 1 Pemodelan dengan seleksi fitur BPSO menggunakan *10-fold cross validation*

## 7. Evaluasi Model

Evaluasi model dilakukan pada set data uji. Pada model dengan seleksi fitur, fitur/kata pada data uji dicocokkan terlebih dahulu dengan fitur terpilih yang dihasilkan dari seleksi fitur BPSO pada data latih. Performa klasifikasi diperoleh dari rata-rata keseluruhan performa klasifikasi pada setiap *fold* pada *10-fold cross-validation*.

## 8. Interpretasi dan penarikan kesimpulan

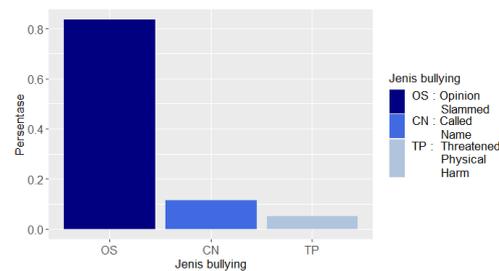
Dikarenakan terdapat 10 kali pengulangan pemodelan pada saat *cross-validation*, maka kombinasi fitur terbaik yang diperoleh pada setiap nilai C sebanyak 10 kombinasi fitur. Interpretasi fitur terpilih secara keseluruhan didasarkan pada fitur yang sering terpilih pada setiap *fold*. Hal ini dilakukan dengan menghitung frekuensi kemunculan

setiap fitur terpilih pada setiap *fold*. Misalkan fitur A muncul sebagai fitur terpilih pada *fold* 1,2,3 dan 4 maka fitur A memiliki frekuensi kemunculan 4 kali. Untuk keperluan interpretasi, digunakan fitur terpilih yang memiliki frekuensi kemunculan terbanyak.

### 3. Hasil dan Pembahasan

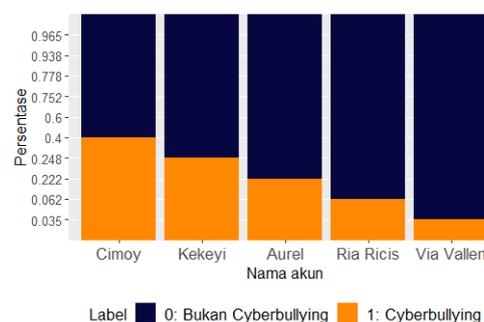
#### 3.1 Eksplorasi Data

Komentar yang termasuk ke dalam *cyberbullying* dikelompokkan menurut jenis *bullying* yang dilakukan apakah termasuk *Called Name* (cn), *Threatened Physical Harm* (tp), atau *Opinion Slammed* (os).



Gambar 2 Grafik perbandingan jenis *bullying*

Berdasarkan Gambar 2 dapat dilihat bahwa jenis *bullying* yang paling banyak dilakukan yaitu berbentuk *Opinion Slammed* (pendapat yang merendahkan) dengan persentase sebanyak 83% dari banyaknya komentar *bullying* yang ada. Kata-kata yang berkaitan dengan hinaan seperti sebutan “tante” mendominasi kategori ini. Terdapat juga kata-kata merendahkan lainnya seperti “menor”, “sensasi”, “plastik”, “gendut”, “jelek”, dan sebagainya. Jenis *bullying* yang paling jarang dilakukan yaitu *Threatened Physical Harm* (mengancam keselamatan fisik) dengan persentase 0,05% dari banyaknya komentar *bullying* yang ada. Kata-kata seperti “bunuh” dan “pergi” mendominasi kategori ini.



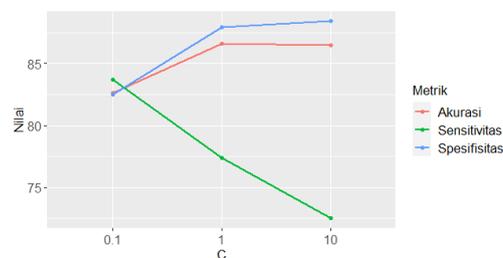
Gambar 3 Proporsi label pada setiap akun

Distribusi jenis label pada setiap akun beraneka ragam. Berdasarkan Gambar 3 dapat dilihat bahwa proporsi komentar yang termasuk *cyberbullying* tidak mendominasi pada setiap akun. Akun yang paling banyak mengalami *cyberbullying* yaitu @cimoyluv dengan persentase komentar *cyberbullying* mencapai 40% dari banyaknya komentar pada akun tersebut. Kata-kata yang berkaitan dengan hinaan fisik wajah dan sebutan negatif berupa nama hewan mendominasi pada akun ini. Akun



### 3.2 Model Support Vector Machine

Proporsi label pada data penelitian ini tidak seimbang yaitu terdiri dari 3830 data (88%) label "0" yang bukan *cyberbullying* dan 539 data (12%) label "1" yang termasuk *cyberbullying*. Ketidakseimbangan data ini akan berdampak pada performa hasil klasifikasi yaitu menghasilkan nilai sensitivitas yang kecil. Sensitivitas menggambarkan seberapa baik model dapat mendeteksi data yang termasuk *cyberbullying* dengan benar. Apabila nilai sensitivitas kecil, maka akan banyak data yang termasuk *cyberbullying* diklasifikasikan secara salah menjadi kelas bukan *cyberbullying* (Somasundaram dan Reddy 2016). Hal ini tentunya tidak diharapkan karena tujuan pemodelan ini kedepannya agar dapat digunakan untuk mendeteksi komentar *cyberbullying* secara otomatis. Apabila banyak komentar *cyberbullying* yang tidak terdeteksi, maka akan menjadi kurang bermanfaat. Untuk itu, perlu dilakukan penanganan ketidakseimbangan data menggunakan SMOTE agar dapat memperbaiki performa klasifikasi dan meningkatkan nilai sensitivitas. Berikut merupakan ringkasan performa klasifikasi pada data uji untuk berbagai nilai C.

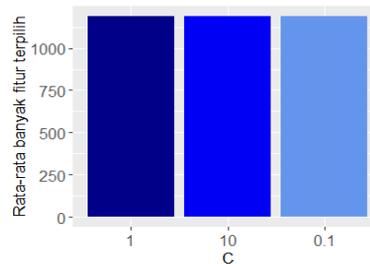


Gambar 6 Performa klasifikasi untuk setiap nilai C

Setelah penerapan SMOTE, nilai sensitivitas pada sebagian besar nilai C terbilang cukup baik yaitu di atas 72%. Nilai sensitivitas menurun seiring bertambahnya nilai C. Sensitivitas tertinggi terdapat pada saat  $C = 0,1$ . Meskipun nilai spesifisitas dan akurasi pada C ini paling rendah diantara C lainnya, akan tetapi nilainya masih baik yaitu di atas 80%. Performa klasifikasi pada semua nilai C terbilang cukup baik, dibuktikan dengan nilai akurasi dan spesifisitas yang dihasilkan di atas 80%. Karena klasifikasi ini berguna untuk mendeteksi *cyberbullying*, maka dibutuhkan nilai sensitivitas yang cukup tinggi tanpa mengabaikan performa lainnya. Maka dari itu, performa terbaik terdapat pada saat  $C = 0,1$  dengan nilai akurasi sebesar 82,6%, sensitivitas 83,7%, dan spesifisitas 82,5%.

### 3.3 Seleksi Fitur BPSO

Setiap kombinasi fitur yang terdapat pada TF-IDF akan dievaluasi menggunakan fungsi *fitness* pada BPSO untuk menentukan banyaknya fitur terbaik. Penerapan BPSO dilakukan dengan bantuan *Library Python Pyswarms*. Nilai parameter yang digunakan pada penelitian ini yaitu  $c_1 = 1,774$ ,  $c_2 = 1,902$ ,  $w = 0,544$ ,  $k = 7$  dengan 10 partikel dan 100 iterasi. Parameter tersebut dipilih setelah dilakukan beberapa percobaan parameter *tuning*.



Gambar 7 Rata-rata banyaknya fitur terpilih untuk setiap nilai C

Dikarenakan terdapat 10 kali pengulangan pemodelan pada saat *cross validation*, maka kombinasi fitur terbaik yang diperoleh pada setiap nilai C sebanyak 10 kombinasi fitur. Gambar 7 menunjukkan rata-rata banyaknya fitur terpilih pada setiap nilai C. Rata-rata banyaknya fitur terpilih pada setiap nilai C tidak terlalu berbeda secara signifikan yaitu berada pada kisaran 1187 – 1191 fitur. Jumlah ini berada pada kisaran 50% dari banyak fitur sebelumnya. Selain dapat mengurangi banyak fitur, penerapan seleksi fitur BPSO juga dapat mengurangi waktu komputasi pada saat pemodelan untuk setiap nilai C.

Tabel 1 Perbandingan performa klasifikasi sebelum dan setelah seleksi fitur BPSO

Nilai C	Sebelum seleksi fitur BPSO			Setelah seleksi fitur BPSO		
	Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)	Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)
0,1	82,6	83,7	82,5	83,1	72,2	84,7
1	86,6	77,4	87,9	87,2	64	90,5
10	86,5	72,5	88,4	86,6	68,3	89,1

Tabel 1 menampilkan nilai rata-rata performa klasifikasi dari 10-fold *cross-validation*. Setelah penerapan seleksi fitur, nilai akurasi dan spesifisitas mengalami kenaikan pada setiap nilai C, sedangkan nilai sensitivitas mengalami penurunan. Nilai sensitivitas mengalami penurunan paling besar pada saat C = 1 yaitu sebesar 13,4%. Nilai spesifisitas mengalami peningkatan lebih dari 2% pada saat C = 1 dan C = 0,1. Peningkatan nilai akurasi pada kedua nilai C ini juga tidak terlalu berbeda jauh. C = 10 mengalami peningkatan nilai akurasi dan spesifisitas yang paling rendah diantara nilai C lainnya. Penerapan seleksi fitur BPSO menghasilkan performa yang baik pada setiap nilai C dengan nilai akurasi dan spesifisitas yang cukup besar. Nilai ini lebih baik daripada model tanpa seleksi fitur. Nilai rata-rata sensitivitas yang diperoleh pada seleksi fitur BPSO masih cukup akurat di atas 60% hanya saja mengalami penurunan daripada model tanpa seleksi fitur.

Gambar 8 menunjukkan nilai performa klasifikasi pada setiap *fold*. Berdasarkan gambar tersebut dapat dilihat bahwa nilai performa klasifikasi pada setiap *fold* beraneka ragam. Terdapat *trade-off* antara nilai sensitivitas dan spesifisitas pada sebagian besar *fold*, yaitu jika nilai spesifisitas besar maka nilai sensitivitas cenderung kecil begitupun sebaliknya. Pada saat C = 0,1, semua nilai performa klasifikasi cenderung mengalami fluktuasi. Terdapat dua *fold* dengan nilai sensitivitas yang cukup besar di atas 83% yaitu pada *fold* ke-5 dan ke-9. Pada saat C = 1, nilai akurasi dan spesifisitas cenderung stabil pada nilai yang cukup tinggi hanya pada dua *fold* terjadi penurunan, sedangkan nilai sensitivitas cenderung fluktuatif pada nilai yang cukup rendah yaitu 48% - 79%. Nilai sensitivitas terbesar diperoleh pada *fold* ke-10 yaitu 79% dengan nilai akurasi dan spesifisitas sebesar 78%. Pada saat C = 10, semua nilai



Gambar 9 menunjukkan kata-kata yang sering muncul berdasarkan fitur terpilih saat  $C = 0,1$ . Fitur-fitur terpilih ini didasarkan pada fitur yang sering terpilih pada setiap *fold*. Berdasarkan gambar tersebut, dapat dilihat bahwa fitur yang sering muncul sebelum seleksi fitur seperti “kakak”, “cantik”, dan “*fallback*” banyak terpilih sebagai fitur pada setiap *fold*. Berdasarkan Gambar 5, kata “kakak” dan “cantik” banyak digunakan baik pada komentar *cyberbullying* maupun bukan *cyberbullying*, sedangkan kata “*fallback*” banyak digunakan pada komentar bukan *cyberbullying*. Fitur yang sering terpilih pada setiap *fold* untuk komentar *cyberbullying* diantaranya yaitu “jelek”, “tante”, “wajah”, “umur”, “pamit”, “alay”, “norak”, dan “menor”. Proporsi fitur pada Gambar 9 cenderung seimbang untuk komentar *cyberbullying* maupun bukan *cyberbullying*.

## 4. Simpulan dan Saran

### 4.1 Simpulan

Pemodelan klasifikasi *cyberbullying* dengan SVM menghasilkan performa klasifikasi yang cukup akurat lebih dari 72% untuk semua performa klasifikasi pada setiap nilai  $C$ . Penerapan seleksi fitur BPSO dapat meningkatkan performa klasifikasi dengan meningkatnya nilai akurasi dan spesifisitas. Akan tetapi untuk kasus pada penelitian ini dipilih model tanpa seleksi fitur pada  $C = 0,1$  karena memiliki nilai sensitivitas paling besar dengan akurasi dan spesifisitas yang cukup baik sehingga dapat mendeteksi komentar *cyberbullying* secara lebih akurat.

### 4.2 Saran

Pada penelitian ini tidak dilakukan konversi emotikon karena data tidak terbaca saat proses *scraping*. Penggunaan metode *scraping* yang lebih baik dapat dilakukan pada penelitian selanjutnya agar data emotikon dapat terbaca dan dilibatkan dalam analisis. Untuk meningkatkan performa klasifikasi pada BPSO, dapat diterapkan modifikasi metode pada BPSO seperti LDIW-PSO, MBPSO, PSO-TVAC, dan sebagainya.

## Daftar Pustaka

- Chrismanto AR, Lukito Y. 2017. Identifikasi komentar spam pada Instagram. *Lontar Komputer*. 8(3): 219-231. Doi: 10.24843/LKJITI.2017.v08.i03.p08.
- Ditch The Label. 2018. 7 facts about *cyberbullying*. [diakses 2020 Des 28]. <https://www.ditchthelabel.org/7-facts-cyberbullying/>.
- Han J, Kamber M, Pei J. 2012. *Data Mining Concepts and Techniques*. Third Edition. USA: Elsevier.
- Hootsuite (*We are Social*). 2020. Digital in 2020. [diakses 2021 Jan 26]. <https://wearesocial.com/digital-2020>.
- Joachims T. 1997. *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*. Germany: Springer.
- Kurniawan M. 2017. Optimasi parameter premise pada adaptive-network-based fuzzy inference system dengan modifikasi kombinasi particle swarm optimization dan genetic algorithm [tesis]. Surabaya: Institut Teknologi Sepuluh November.

- Laia ML, Setyawan Y. 2020. Perbandingan hasil klasifikasi curah hujan menggunakan metode svm dan nbc. *Jurnal Statistika Industri dan Komputasi*. 5(2): 51-61.
- Luqyana WA, Cholissodin I, Perdana RS. 2018. Analisis sentimen *cyberbullying* pada komentar Instagram dengan metode klasifikasi *support vector machine*. *J-PTIIK*. 2(11): 4704-4713.
- Mafarja M, Jarrar R, Ahmad S, Abusnaina AA. 2018. Feature selection using binary particle swarm optimization with time varying inertia weight strategies. *Proceedings of the 2<sup>nd</sup> International Conference on Future Networks and Distributed Systems*: 1-9.
- Maulida I, Suyatno A, Hatta HR. 2016. Seleksi fitur pada dokumen abstrak teks Bahasa Indonesia menggunakan metode *information gain*. *JSM STMIK Mikroskil*. 17(2):249-258.
- Pajri D, Umaidah Y, Padilah TN. 2020. K-nearest neighbor berbasis particle swarm optimization untuk analisis sentimen terhadap Tokopedia. *Jurnal Teknik Informatika dan Sistem Informasi*. 6(2):242-253.
- Rifauddin M. 2016. Fenomena *cyberbullying* pada remaja: studi analisis media sosial facebook. *Jurnal Ilmu Perpustakaan, Informasi, dan Kearsipan Khizanah Al-Hikmah*. 4(1):35-44.
- Somasundaram A, Reddy US. 2016. Data imbalance: Effects and solutions for classification of large and highly imbalanced data. *ICRECT*. 28-34.
- Vieira SM, Mendonca LF, Farinha GJ, Sousa JM. 2013. Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. *Applied Soft Computing*. 13(8): 3494-3504.
- Widiyanto MT. 2018. Optimasi PSO untuk metode clustering fuzzy c-means dalam pengelompokan kelas. *Jurnal PETIR*. 11(1): 72-91.
- Yan HE, Wei JMA, Ji PZ. 2016. The parameters selection of PSO algorithm influencing on performance of fault diagnosis. *MMME*. 63.