Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 5 Nomor 1, Maret 2021, pp. 10-19
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

# Deteksi *Cyberbullying* berdasarkan Unsur Perbuatan Pidana yang Dilanggar dengan *Naive Bayes* dan *Support Vector Machine*

**Tommy Nugraha Manoppo[1], Dhomas Hatta Fudholi[2]**
[1,2]Program Studi Informatika Program Magister, Universitas Islam Indonesia
Jl.Kaliurang, KM. 14.5, Sleman, D.I. Yogyakarta, 55584
17917224@students.uii.ac.id[1], hatta.fudholi@uii.ac.id[2]

***Abstract***

*Lack of understanding by Indonesian social media user about law impact inflicted to cyberbullying perpetrators makes many cyberbullying cases has not handled properly and ended up with nothing. Indonesia hasn't yet law authority that govern cyberbullying in specific, causing no guideline regard the definition about cyberbullying itself. There is an extension about definition of violence which state that violence is not only physically deliver, but also psychologically, referred an inferences cyberbullying characteristics possibly qualify in element of criminal act. Therefore, the element of criminal act can be used as a basis for detecting potential of cyberbullying. In this research, literature review is used to determine the elements of criminal acts related to the characteristics of cyberbullying and also in finding a model classifier to detect cyberbullying messages. So there are 5 criminal acts related to cyberbullying characteristic which insult, accuse with defamation, hatred about ethnicity, religion, race and inter-group relations, threat of violence, and threat of telling secret. Total of 5000 tweets are collected as a dataset. Feature extraction, using the N-gram method with TF-IDF weighting is expected to obtain sentiment based on the use of words. The context of language becomes important in this study, so the dataset annotation process is carried out by linguist. The results on the application of the two model classfier were Naïve Bayes and SVM after applying resampling by over-sampling using SMOTE method, can correctly predict the potential for cyberbullying by their violated element of criminal act with the average performance measurement of 90%.*

***Keywords****: cyberbullying, elements of criminal act, detection*

***Abstrak***

*Pemahaman pengguna media sosial khususnya di Indonesia yang kurang terhadap jerat hukum kepada pelaku cyberbullying, membuat banyak kasus cyberbullying tidak tertangani secara serius. Belum adanya kebijakan hukum yang mengatur secara rinci cyberbullying, menyebabkan tidak adanya pedoman mengenai definisi cyberbullying itu sendiri. Perluasan definisi kekerasan yang tidak hanya berlaku secara fisik, tapi juga psikis, memberikan kesimpulan bahwa karakteristik cyberbullying dapat digunakan untuk memenuhi unsur perbuatan pidana. Maka dari itu, unsur perbuatan pidana dapat menjadi landasan dalam melakukan deteksi cyberbullying. Pada penelitian ini, studi literatur digunakan untuk menentukan unsur perbuatan pidana terkait karakteristik cyberbullying dan juga dalam menemukan model classifier dalam proses klasifikasi pesan cyberbullying. Disimpulkan ada 5 unsur perbuatan pidana terkait karakteristik cyberbullying yaitu penghinaan, menuduh dan bersifat pencemaran, rasa kebencian terkait SARA, ancaman kekerasan, dan ancaman membuka rahasia. Sebanyak 5000 tweet dikumpulkan sebagai dataset. Ekstraksi fitur, menggunakan metode N-grams dengan pembobotan TF-IDF yang diharapkan dapat memperoleh sentiment berdasarkan penggunaan bahasa, sehingga model classifier memiliki ruang pengetahuan dalam mengenali pesan yang berpotensi cyberbullying. Konteks bahasa menjadi penting dalam hal ini, sehingga proses anotasi dataset dilakukan oleh seorang ahli linguistik. Hasilnya pada penerapan dua model klasifikasi Naïve Bayes dan Support Vector Machine setelah dilakukan resampling dengan over-sampling menggunakan metode SMOTE,*

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 5 Nomor 1, Maret 2021, pp. 10-19
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

*mampu memprediksi benar potensi cyberbullying berdasarkan unsur perbuatan yang dilanggar dengan performance measurement rata-rata diatas 90%.*

***Kata kunci**: cyberbullying, unsur perbuatan pidana, deteksi*

## 1. PENDAHULUAN

Media sosial memungkinkan penggunanya untuk dapat berkomunikasi secara interaktif, saling bertukar opini sebagai bentuk komunikasi yang ekspresif dengan pengguna lainnya. Tentunya, karena media sosial juga dianggap sebagai ruang publik, maka penggunanya dapat bebas dalam bertukar pandangan dan pendapat sebagai aktualisasi terhadap *freedom of expression* dasar dalam menjamin hak asasi manusia [1]. Alasan *freedom of expression* dalam "melemparkan" opini yang tidak terkontrol di media sosial, berpotensi memberikan dampak negatif karena perpesktif pengguna yang bisa berbeda-beda dalam merespon opini. Pemahaman yang kurang mengenai aturan sosial juga seharusnya berlaku di media sosial dan juga fakta bahwa banyak pengguna yang mengabaikan *privacy policy* (PP) dan *terms of service* (TOS) berisi aturan dan prosedur penggunaan media sosial [2], menyebabkan pengguna menganggap tidak ada aturan dan batasan dalam beropini di media sosial, membuat pengguna menjadi tidak respek terhadap perbedaan pendapat dan berakhir melontarkan respon negatif. Respon negatif di media sosial yang bersifat menghina, kasar, vulgar yang spesifik ditujukan pada seseorang atau kelompok dapat dikenali sebagai tindakan *cyberbullying* [3].

Hasil *survey* asosiasi penyelenggara jasa internet Indonesia tahun 2019 melaporkan bahwa 49% pegguna internet di Indonesia pernah mengalami atau menerima pesan *cyberbullying* di media sosial, dan dari 49% diantaranya, 31.6% membiarkan tanpa adanya tindak-lanjut, sementara hanya 3.6% kasus yang dapat diselesaikan melalui jalur hukum [4]. Hal ini dapat dikarenakan masih kurangnya pemahaman pengguna media sosial terhadap jerat hukum yang dapat dikenai kepada pelaku *cyberbullying*. Mendeteksi indikasi potensi *cyberbullying* dalam penggunaan media sosial dapat menjadi langkah awal dalam penanganan kasus *cyberbullying*.

Deteksi terhadap pesan yang terindikasi *cyberbullying* masih menjadi tantangan, khususnya menyangkut data media sosial di Indonesia. Karena nyatanya, belum ada kebijakan hukum terhadap kasus *cyberbullying* yang diatur secara rinci dalam peraturan perundang – undangan di Indonesia sampai saat penelitian ini dibuat, sehingga tidak ada pedoman definisi yang jelas mengenai *cyberbullying* itu sendiri.

Kasus *cyberbullying* yang sering terjadi di Indonesia cenderung hanya dianggap sebagai suatu tindakan yang melanggar unsur perbuatan penghinaan [5], padahal beberapa literasi [3], [6], [7] mengemukakan bahwa tindakan *cyberbullying* memiliki berbagai jenis karakteristik. Di Indonesia regulasi hukum terkait dengan ruang siber diatur dalam Undang-Undang Nomor 19 Tahun 2016 perubahan atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik, dimana perbuatan pada

pasal utamanya notabene digolongkan sebagai suatu tindak pidana yang-mana juga termasuk pasal mengenai penghinaan. Secara lebih umum suatu tindak pidana diatur dalam Kitab Undang-Undang Hukum Pidana (KUHP). Maka dari itu, pada penelitian ini studi literatur digunakan untuk menemukan dan menentukan unsur perbuatan pidana apa saja yang terkait dengan karakteristik *cyberbullying*, sehingga unsur perbuatan tersebut dapat menjadi landasan yang kuat dalam melakukan identifikasi atau mengenali suatu tindakan *cyberbullying* yang terjadi di media sosial Indonesia.

Deteksi *cyberbullying* dapat dilakukan secara otomatis dengan beberapa model *classifier* pada *data mining* berdasarkan karakteristik isi pesan [8]. Hal ini bertujuan sebagai respon cepat dalam penanganan kasus *cyberbullying* di media sosial. Beberapa model *classifier* yang telah digunakan pada penelitian terdahulu dalam mendeteksi pesan *cyberbullying*, menunjukkan akurasi yang tinggi, seperti penerapan Support Vector Machine (SVM), K-Nearest Neighbor (KNN), dan Naïve Bayes, yang memberikan rata-rata akurasi diatas 90% [9]. Pada penelitian ini, model Naïve Bayes dan SVM digunakan dengan mempertimbangkan bahwa model tersebut *suitable* terhadap pengolahan data yang besar, mengingat fitur yang dibangun menggunakan metode N-grams yang akan membuat *text vectorization* dari perpaduan kata atau *token* dengan rentang sebanyak *n* kata yang ada dalam *dataset* sehingga menghasilkan fitur yang jumlahnya besar. Meskipun demikian, N-grams tetap dipilih untuk digunakan karena dapat memahami pesan lebih baik dibandingkan Bag-of-Words yang membuat fitur hanya dari kata-perkata [10]. Misalnya untuk menentukan *sentiment* negatif pada suatu pesan, “tidak suka” dari fitur 2-gram atau *bigram* lebih mudah dinilai dari pada “tidak” dan “suka” jika menggunakan Bag-of-Words. Fitur yang terbentuk dengan N-grams kemudian dibobotkan dengan perhitungan Terms Frequency – Inverse Document Frequency (TF-IDF).

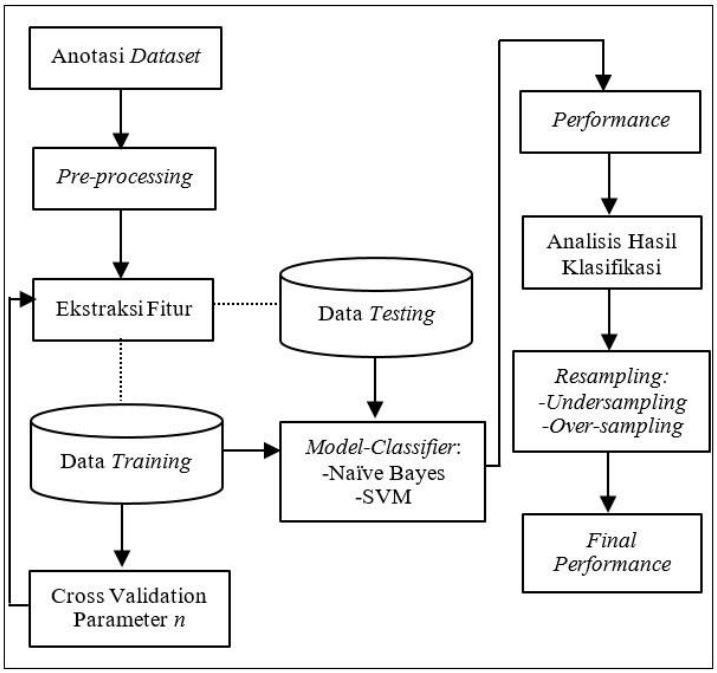
Mengaitkan ketiga aspek pada penelitian ini, yakni hukum, perspektif Bahasa, dan teknik *data-mining* diharapkan dapat memberikan hasil yang valid secara metodologi ilmiah dalam melakukan deteksi otomatis pesan *cyberbullying* di media sosial Indonesia. Berbeda dengan penelitian serupa, untuk menciptakan hasil klasifikasi yang valid, maka dalam proses anotasi, dilakukan kerjasama penelitian dengan seorang ahli linguistik Bahasa Indonesia untuk mengenali pesan yang berpotensi sebagai *cyberbullying* berdasarkan unsur perbuatan yang dilanggar berdasarkan pendekatan ilmu linguistik. Selain itu, pada penelitian ini, identifikasi pesan berpotensi *cyberbullying* dikenali berdasarkan unsur perbuatan yang dilanggar. Unsur perbuatan dapat menjadi landasan untuk mengetahui karakteristik *cyberbullying*. Unsur perbuatan yang dimaksud adalah aspek perbuatan melanggar hukum yang dapat dipertanggung-jawabkan secara pidana terkait dengan karakteristik *cyberbullying* yang diatur dalam Undang-Undang. Pada penelitian ini, potensi *cyberbullying* akan diidentifikasi berdasarkan 5 unsur perbuatan yang diinisialisasikan sebagai unsur Q: Menyerang kehormatan dengan menghina (penghinaan) tidak bersifat pencemaran, unsur W:

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 5 Nomor 1, Maret 2021, pp. 10-19
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

Menyerang kehormatan atau nama baik dengan menuduh dan bersifat pencemaran, unsur E: Menimbulkan rasa kebencian atau permusuhan menyangkut Suku, Agama, Ras, dan Antargolongan tertentu, unsur R: Ancaman kekerasan atau menakut-nakuti, dan unsur T: Ancaman membuka rahasia.

## 2. METODOLOGI PENELITIAN

Rangkaian proses penelitian dimulai dengan mengumpulkan *dataset* berupa *tweet*. *Dataset* kemudian dilakukan anotasi atau *data labelling* seperti alur penelitian pada Gambar 1.




**Gambar 1.** Alur Penelitian

Anotasi bertujuan untuk membangun *knowledge* pada model *classifier* [11] dan juga digunakan untuk validasi hasil klasifikasi. Anotasi membutuhkan annotator, yakni orang yang berkapasitas untuk melakukan anotasi sesuai dengan keperluan penelitian. Kapasitas annotator menjadi salah satu penentu model penelitian menjadi valid dan tidak mengalami kesalahan metodologi, maka annotator pada penelitian ini merupakan seorang ahli linguistik, sehingga identifikasi kasus *cyberbullying* pada *dataset* menggunakan pendekatan metode linguistik. Annotator bertugas untuk mengenali setiap data berdasarkan potensinya sebagai suatu tindak *cyberbullying* sesuai dengan unsur perbuatan yang dilanggar. *Dataset* hasil anotasi kemudian akan melewati tahap *pre-processing*. Data yang diperoleh memuat banyak *noise*, seperti tanda *mention*, *hashtag*, tanda RT atau *retweet* yang biasanya merupakan tipikal dari *tweet*. Konten seperti itu notabene tidak digunakan dalam proses klasifikasi, maka *noise* perlu dihilangkan untuk

mengurangi *cost* komputasi dan mengurangi resiko terjadinya kesalahan klasifikasi akibat *noise* tersebut mengingat bahwa seharusnya data yang diolah berfokus pada penggunaan struktur bahasa seperti paduan kata ataupun penjelesan yang berusaha diutarakan pada suatu kalimat. Setelah menghilangkan *noise*, tahap *pre-processing* berikutnya adalah merekontruksi seluruh penggunaan huruf menjadi huruf kecil atau biasa disebut *case folding*, tujuannya agar kata yang sama nantinya tidak dianggap sebagai fitur yang berbeda hanya karena perbedaan penggunaan huruf kapital. Langkah terakhir pada tahap *pre-processing* adalah melakukan *tokenizing* atau setiap data dipisah menjadi kata-perkata berdasarkan *delimeter* spasi untuk menghasilkan *n token* berupa keseluruhan kata yang ada dalam *dataset*.

Pada penelitian ini, *token* hasil *pre-processing* akan digunakan dalam tahap ekstraksi fitur menggunakan metode N-grams. Metode ekstraksi menggunakan N-grams akan menghasilkan fitur berupa perpaduan kata atau *token* yang ada dalam *dataset* berdasarkan nilai parameter *n*, maka untuk mengetahui penggunaan nilai parameter dengan nilai *n* berapa yang paling optimal untuk membangun fitur digunakan Cross Validation pada data *training*. Setiap percobaan hasil Cross Validation untuk nilai parameter *n* pada model Naïve Bayes dan SVM akan dibandingkan hasilnya berdasarkan *performance measurement* berupa Positive Predictive Value (PPV), Negative Predictive Value (NPV), *sensitivity*, dan *specificity*. Hasil performa terbaik terhadap nilai parameter *n* yang akan digunakan membangun ulang fitur.

Tahapan berikutnya kemudian dilakukan klasifikasi berdasarkan berdasarkan unsur perbuatan yang dilanggar. Ketika suatu pesan dianggap telah memiliki suatu unsur perbuatan, maka secara otomatis pesan tersebut juga memenuhi karakteristik *cyberbullying*. Misalnya suatu pesan, setelah dianalisis ternyata memenuhi unsur perbuatan penghinaan tanpa pencemaran, maka pesan tersebut memiliki karakteristik *cyberbullying*, karena menghina merupakan suatu karakteristik *cyberbullying* dengan jenis *harassment* [3], [6], [7]. Setiap unsur perbuatan tidak mempengaruhi satu-sama lain, dalam artian suatu data yang terindikasi *cyberbullying* bisa memiliki beberapa unsur perbuatan sekaligus. Maka dari itu, proses klasifikasi harus dilakukan terpisah untuk setiap unsur perbuatan, karenanya, akan terdapat 5 kali proses klasifikasi dengan data uji yang sama dengan kelima label unsur perbuatan untuk dapat mendeteksi seluruh unsur perbuatan.

Dari 2110 data, faktanya kasus *cyberbullying* yang terjadi di media sosial Indonesia sesuai dengan sampel *dataset* yang diperoleh dari Twitter pada rentang waktu tertentu paling banyak melanggar unsur perbuatan Q sebanyak 2044 kasus, dan paling jarang ditemukan kasus yang melanggar unsur perbuatan T, yakni hanya 4 kasus. Fakta tersebut mengindikasikan bahwa suatu kasus *cyberbullying* belum tentu melanggar seluruh unsur perbuatan, sehingga hal ini menyebabkan terjadinya *imbalanced* pada distribusi data masing-masing kelas setiap unsur perbuatan yang membuat jumlah data untuk masing-masing kelas tidak merata (contohnya data yang

memenuhi unsur perbuatan Q sangat banyak dibandingkan dengan yang tidak memenuhi unsur perbuatan Q). Keadaan *dataset* yang *imbalanced* menyebabkan model tidak dapat membangun dengan baik *knowledge* untuk bisa memprediksi kelas minoritas, sehingga model akan lebih dominan memprediksi kelas minoritas sebagai kasus kelas mayoritas. Salah satu solusi menangani *dataset imbalanced* adalah dengan melakukan *resampling* terhadap dataset.

### 2.2. Resampling Menggunakan RUS dan SMOTE

Distribusi kelas yang tidak merata dimana kelas yang memiliki data yang lebih banyak menjadi mayoritas dan data pada kelas yang lebih sedikit menjadi minoritas dan inilah yang disebut sebagai *class imbalanced* atau *dataset imbalanced* [13]. Pada kasus penelitian ini, permasalahan imbalanced terjadi terhadap masing – masing dataset unsur perbuatan. Kondisi ini membuat pengetahuan pada kelas minoritas terlalu sedikit untuk dipelajari oleh model menyebabkan model tidak dapat secara efektif dan benar memprediksi kasus *cyberbullying*.

Solusi untuk menangani *imbalanced dataset* dapat dilakukan dengan melakukan *resampling* terhadap *dataset*. Terdapat dua jenis metode *resampling* secara umum, yaitu *resamping* dengan *under-sampling* dan *resampling* dengan *oversampling* [12]. Metode *resampling* dapat dilakukan dengan cara sederhana menggunakan *under-sampling* atau *down-sampling*, yaitu metode untuk melakukan seleksi terhadap data pada kelas mayoritas sedemikian sehingga menghasilkan *subset* data yang seimbang dengan data pada kelas minoritas [14]. Salah satu metode *under-sampling* adalah Random Under-sampling (RUS).

RUS secara sederhana memilih secara acak data pada kelas mayoritas sebanyak jumlah data pada kelas minoritas. Selain *under-sampling* metode *resampling* yang digunakan lainnya adalah *over-sampling* dengan SMOTE. Jika RUS mengurangi jumlah data pada kelas mayoritas, pada SMOTE justru menambahkan data secara sintetis pada kelas minoritas sehingga jumlahnya seimbang dengan data pada kelas mayoritas. Beberapa penelitian terkait dengan penerapan SMOTE untuk menangani masalah *imbalanced dataset* diketahui secara signifikan dapat meningkatkan performa model pada *low dimension dataset* dan juga pada beberapa kasus *high dimension dataset* [15], [16].

Namun terdapat penelitian yang memberikan kesimpulan bahwa performa yang baik terhadap model klasifikasi dalam menangani *imbalanced dataset* dapat dicapai tanpa perlu adanya banyak perubahan pada *dataset* asli, yaitu seperti dengan hanya menerapkan RUS [17]. Maka dari itu, pada penelitian ini kedua metode tersebut digunakan dan kemudian perlu dibandingkan kinerjanya dalam menangani *imbalanced dataset* pada kasus klasifikasi pesan *cyberbullying*.

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 5 Nomor 1, Maret 2021, pp. 10-19
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

## 3. HASIL DAN PEMBAHASAN

### 3.1. Hasil *Cross Validation* Parameter *n* Pada Naive Bayes dan SVM

Hasil implementasi terhadap variatif nilai parameter *n* dengan *model classifier* Naïve Bayes dan SVM secara berturut-turut terhadap 2-gram, 3-gram, 4-gram, hingga 5-gram seperti pada Tabel 1 menunjukkan bahwa semakin meningkatnya nilai *n* dalam artian semakin banyak kombinasi kata dari setiap fitur yang digunakan, membuat *model classifier* akan cenderung mengklasifikasikan kasus kepada suatu kelas, dan membuat performa klasifikasi terhadap *model classifier* tidak mampu memprediksi dengan baik kasus *kelas lainnya*. Hal ini ditunjukkan dengan semakin menurunnya performa tertentu pada *classifier* jika nilai *n* meningkat. Maka disimpulkan, penggunaan 2-gram memiliki performa yang paling optimal untuk digunakan dalam membangun ulang fitur.

**Tabel 1.** Hasil *Cross Validation* Parameter *n*

| Parameter Nilai *n* | *Naïve Bayes (%)* | | | | *SVM(%)* | | | |
|---|---|---|---|---|---|---|---|---|
| | **PPV** | **NPV** | **Sens.** | **Spec.** | **PPV** | **NPV** | **Sens.** | **Spec.** |
| 2-gram | 69.24 | 82.06 | 77.59 | 74.84 | 87.19 | 71.39 | 47.94 | 94.86 |
| 3-gram | 63.22 | 84.30 | 83.55 | 64.51 | 93.84 | 64.88 | 26.81 | 98.71 |
| 4-gram | 47.55 | 87.48 | 95.46 | 23.13 | 96.97 | 61.08 | 13.00 | 99.70 |
| 5-gram | 45.84 | 87.08 | 96.61 | 16.66 | 96.70 | 60.79 | 11.92 | 99.70 |

*PPV=Positive Predictive Value, NPV=Negative Predictive Value, Sens.=Sensitivity, Spec.=Specificity.*

### 3.2. Hasil Klasifikasi Sebelum dan Sesudah *Under-Sampling*

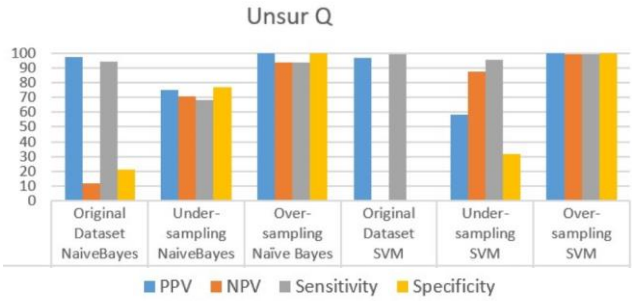
*Under-sampling* dilakukan dengan membuat jumlah data untuk setiap kelas pada setiap unsur perbuatan menjadi relatif sama. Maka dari itu, agar jumlah data sama, maka jumlah data pada kelas mayoritas harus diminimalkan sehingga sama jumlahnya dengan maksimal jumlah data pada kelas minoritas. Meminimalkan data pada kelas mayoritas dilakukan secara *random* atau *shuffled sampling*. *Dataset* hasil *under-sampling* kemudian diuji menggunakan Cross-Validation terhadap kedua *model classifier*.

Hasil pada Gambar 2 menunjukkan performa yang memperhitungkan kelas minoritas, yaitu jika kondisi kelas minoritas ada pada kelas negatif seperti kasus unsur perbuatan Q, maka performa NPV dan *specificity* sebelum *under-sampling* rendah akan secara konsisten meningkat setelah *under-sampling*, dan jika kondisi kelas minoritas ada pada kelas positif seperti kasus unsur perbuatan W,E,R, dan T pada Gambar 3, Gambar 4, Gambar 5, dan Gambar 6, maka performa PPV dan *sensitivity* yang sebelum *under-sampling* rendah juga sebagian besar secara konsisten meningkat setelah *under-sampling* baik pada kedua *classifier*, kecuali performa unsur perbuatan E pada *model classifier* SVM mengalami penurunan terhadap *sensitivity* yang-mana memperhitungkan kasus kesalahan pada kelas minoritas.

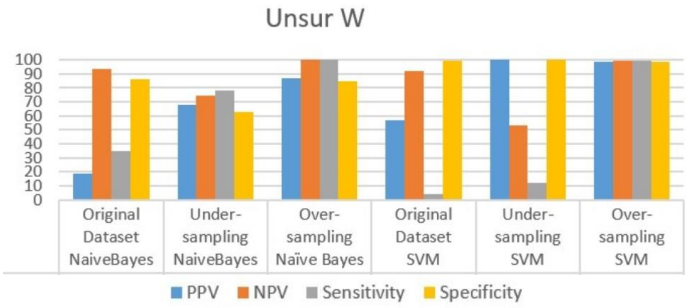
Sementara pada hasil klasifikasi pada *dataset* dengan porsi yang telah mengalami *over-sampling* menggunakan SMOTE memberikan performa yang secara konsisten sangat baik untuk kedua *classifier*, sehingga hasil performa dengan *over-sampling* lebih tinggi dibandingkan dengan *under-sampling*

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 5 Nomor 1, Maret 2021, pp. 10-19
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

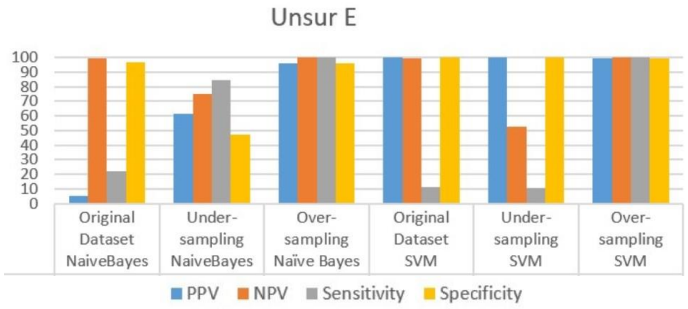
sebelumnya. Jika pada *dataset* dengan kondisi *under-sampling* berhasil menaikkan performa yang memperhitungkan kelas minoritasnya tetapi cenderung menurunkan performa yang memperhitungkan kelas mayoritas, sementara pada kondisi *dataset* yang telah mengalami *over-sampling*, bukan hanya menaikkan performa yang memperhitungkan kelas minoritasnya saja, namun juga performa yang memperhitungkan kelas mayoritas juga tetap tinggi.



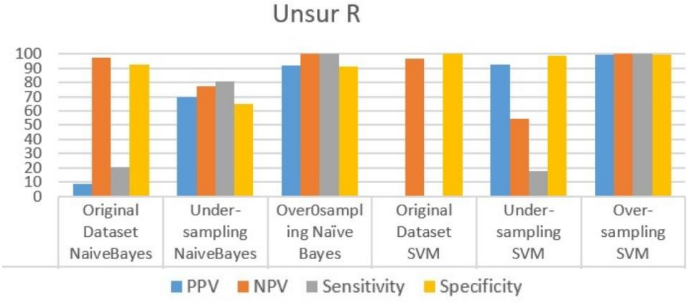

**Gambar 2.** Hasil Klasifikasi Unsur Q



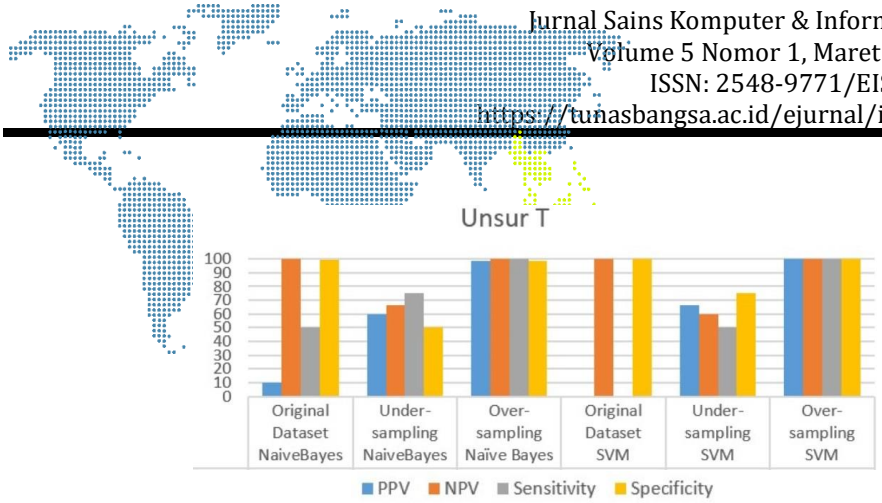

**Gambar 3.** Hasil Klasifikasi Unsur W




**Gambar 4.** Hasil Klasifikasi Unsur E




**Gambar 5.** Hasil Klasifikasi Unsur R




**Gambar 6.** Hasil Klasifikasi Unsur T

## 4. SIMPULAN

Pesan *cyberbullying* dapat dikenali dari berbagai sisi perspektif isi pesannya. Pada penelitian ini, identifikasi pesan *cyberbullying* dikenali berdasarkan unsur perbuatan yang dilanggar sesuai karakteristik *cyberbullying*-nya dengan menggunakan model klasifikasi Naïve Bayes dan SVM. Hasil klasifikasi menunjukkan bahwa implementasi model Naïve Bayes sebelum dilakukan resampling lebih handal dibandingkan SVM dengan pertimbangan kasus *false negative* lebih banyak terjadi pada SVM sehingga performa *sensitivity* pada SVM menjadi sangat rendah.

Performa klasifikasi terhadap dataset yang imbalanced mengalami peningkatan setelah dilakukan *resampling* terhadap *dataset* baik dengan under-sampling ataupun over-sampling. Penerapan *over-sampling* menggunakan metode *syntetic* seperti SMOTE pada model SVM menunjukkan performa paling baik dengan *rate* rata-rata diatas 90%. Hal tersebut memberikan kesimpulan bahwa klasifikasi pesan *cyberbullying* dengan mengaitkan karakteristik *cyberbullying* yang dikaji berdasarkan kebijakan hukum berdasarkan unsur perbuatan yang dilanggar dan perspektif bahasa yang dilakukan pada penelitian ini dapat memenuhi identifikasi pesan *cyberbullying* di media sosial Indonesia.

## DAFTAR PUSTAKA

[1] J. Wilson and N. Gapsiso, 2014, "Social Media and the Freedom of Expression in Nigeria : Posting the mind of a Nation," *Int. J. Internet Trolling Online Particip. Soc.*, vol. 1, no. 1, pp. 5–22.

[2] J. A. Obar and A. Oeldorf-Hirsch, 2016, "The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services."

[3] P. J. Larson, *Safe And Sound : Social Media*, 2nd ed. Huntington Beach, CA: Cracchiolo, Rachelle, 2017.

[4] APJII, "Survei APJII: 49% Pengguna Internet Pernah Dirisak di Medsos," 2019. [Online]. Available: https://databoks.katadata.co.id/datapublish/2019/05/16/survei-apjii-49-pengguna-internet-pernah-dirisak-di-medsos.

[5] E. N. Putra, "Merunut Lemahnya Hukum Cyberbullying di Indonesia,"

2019. https://theconversation.com/merunut-lemahnya-hukum-cyberbullying-di-indonesia-110097.

[6] N. Willard, 2007, "Effectively Managing Internet Use Risks in Schools," *Online*, pp. 1–19.

[7] S. Chadwick, 2014, "Impacts of Cyberbullying, Building Social and Emotional Resilience in Schools," p. 89, doi: 10.1007/978-3-319-04031-8.

[8] H. Rosa *et al.*, 2019, "Automatic cyberbullying detection: A systematic review," *Comput. Human Behav.*, vol. 93, pp. 333–345, doi: 10.1016/j.chb.2018.12.021.

[9] N. Abdulloh and A. Fathan, 2019, "Deteksi Cyberbullying pada Cuitan Media Sosial Twitter," vol. 01.

[10] B. Li, T. Liu, Z. Zhao, P. Wang, and X. Du, 2017, "Neural bag-of-ngrams," *31st AAAI Conf. Artif. Intell. AAAI 2017*, pp. 3067–3074.

[11] A. Wang, "What Is Data Annotation?," 2019. https://www.quora.com/What-is-data-annotation.

[12] J. Brownlee, "8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset," 2015. https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/ (accessed Jul. 27, 2020).

[13] D. Ramyachitra and P. Manikandan, 2014, "Imbalanced dataset classification and solutions: a review," *Int. J. Comput. Bus. Res.*, vol. 5, no. 4.

[14] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, 2004, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, doi: 10.1145/1007730.1007735.

[15] S. Maldonado, J. López, and C. Vairetti, 2019, "An alternative SMOTE oversampling strategy for high-dimensional datasets," *Appl. Soft Comput. J.*, vol. 76, pp. 380–389, doi: 10.1016/j.asoc.2018.12.024.

[16] D. Elreedy and A. F. Atiya, 2019, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," *Inf. Sci. (Ny).*, vol. 505, pp. 32–64, doi: 10.1016/j.ins.2019.07.070.

[17] T. Hasanin and T. M. Khoshgoftaar, 2018, "The effects of random undersampling with simulated class imbalance for big data," *Proc. - 2018 IEEE 19th Int. Conf. Inf. Reuse Integr. Data Sci. IRI 2018*, pp. 70–79, doi: 10.1109/IRI.2018.00018.