

Deteksi Komentar Cyberbullying Pada YouTube Dengan Metode Convolutional Neural Network – Long Short-Term Memory Network (CNN-LSTM)

Albertus Josef Andika^{1*}, Yosi Kristian², Esther Irawati Setiawan³

^{1,2,3} Magister Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur
Email: ^{1*} albertusja01@gmail.com, ² yosi@stts.edu, ³ esther@stts.edu

(Naskah masuk: 24 Jul 2023, direvisi: 11 Sep 2023, diterima: 27 Sep 2023)

Abstrak

Pada era digital seperti sekarang *cyberbullying* kerap kali terjadi di berbagai belahan dunia termasuk di Indonesia, hal ini dapat terjadi pada siapa saja dan dimana saja terutama media sosial seperti YouTube melalui fitur komentar semua pengguna yang memiliki akun dapat dengan mudah terlibat *cyberbullying* hanya melalui berbalas komentar. Penelitian ini bertujuan untuk melakukan deteksi adanya *cyberbullying* melalui pengumpulan serta pengklasifikasian komentar negatif video pada kanal YouTube dengan konten tertentu berbasis bahasa Indonesia (serta bahasa-bahasa daerah tertentu, seperti Jawa dan Surabaya) melalui metode *deep-learning Convolutional Neural Network – Long Short-Term Memory Network (CNN-LSTM)*. *Dataset* komentar yang dipakai dalam penelitian dikumpulkan dengan menggunakan *Application Program Interface (API)* yang telah disediakan oleh Youtube secara gratis dan terbatas kuota secara kumulatif. Terkumpul data komentar total sebanyak 26.918 komentar dengan rincian 9.834 komentar terklasifikasi *cyberbullying* dan 17.084 komentar terklasifikasi sebagai bukan *cyberbullying*. Setelah *dataset* dipakai dalam proses *training* pada model CNN-LSTM dan menghasilkan sebuah model dengan nilai *F1-score* sebesar 0,84, model tersebut dipakai dalam sebuah API sederhana yang menerima *input* beberapa kalimat yang akan dideteksi konten *cyberbullying* dan menghasilkan *output* berupa JSON yang berisi hasil klasifikasi dari setiap kalimat yang akan dideteksi.

Kata Kunci: *Cyberbullying Detection, CNN-LSTM, Youtube, Scrapping, Dataset Indonesia.*

Detection of Cyberbullying Comments on Youtube Social Media Using Convolutional Neural Network – Long Short Term Memory Network (CNN-LSTM) Method

Abstract

In the digital era, cyberbullying often occurs in various parts of the world, including Indonesia. This can happen to anyone and anywhere, especially on social media sites such as YouTube. Through the comments feature, all users who have an account can easily engage in cyberbullying simply by responding to comments. This study aims to detect cyberbullying by collecting and classifying negative video comments on YouTube channels with certain content based on Indonesian (as well as certain regional languages, such as Javanese and Surabaya) through the Convolutional Neural Network Long Short-Term Memory Network (CNN-LSTM) deep-learning method. The comment dataset used in the research was collected using the Application Program Interface (API), which has been provided by Youtube free of charge and with a cumulatively limited quota. Comment data collected totaled 26,918 comments, with details of 9,834 comments classified as cyberbullying and 17,084 comments classified as not cyberbullying. After the dataset is used in the training process on the CNN-LSTM model and produces a model with an F1-score of 0.84, the model is used in a simple API that accepts input of a few sentences that will detect cyberbullying content and produces output in the form of JSON, which contains the classification results of each sentence to be detected.

Keywords: *Cyberbullying Detection, CNN-LSTM, Youtube, Scrapping, Indonesian Dataset*

I. PENDAHULUAN

Komunikasi interpersonal adalah komunikasi yang terjadi antar dua orang atau lebih melalui berbagai macam media baik secara langsung maupun tidak langsung. Seiring dengan dunia sekarang, jenis telekomunikasi pun juga meningkat hal ini dapat terlihat jelas melalui perkembangan teknologi yang semakin canggih dan cepat sehingga dapat mempertemukan dua orang dari belahan dunia yang berbeda pada satu waktu yang sama. Namun seiring dengan berjalannya waktu komunikasi terutama yang berbeda kultur atau budaya seperti di Indonesia juga kerap kali menyebabkan timbulnya perbedaan pendapat atau miss komunikasi yang menyebabkan pergeseran makna dan kesalahpahaman. Berbalas komentar bisa bernada saling menyerang serta memicu permusuhan yang biasa disebut dengan istilah perundungan atau *bullying*.

Pada era sekarang *bullying* tidak hanya terjadi di dunia nyata namun juga sangat mungkin dan sering terjadi di dunia maya terutama dipicu karena hal sepele seperti perbedaan pendapat dan pergeseran makna dari sebuah komentar pada suatu unggahan di media sosial.

Cyberbullying adalah perlakuan yang ditujukan untuk mempermalukan, menakut-nakuti, melukai, atau menyebabkan kerugian bagi pihak yang lemah dengan menggunakan sarana komunikasi Teknologi Informasi [1]. *Cyberbullying* dapat juga diartikan sebagai penggunaan teknologi informasi dan komunikasi oleh individu maupun kelompok untuk melecehkan seseorang [2].

Sebuah penelitian yang dilakukan oleh lembaga riset *cyberbullying* memaparkan bahwa jumlah kasus *cyberbullying* setiap tahun cenderung mengalami peningkatan seperti terlihat pada Gambar 1 [3].

Salah satu media sosial berbasis video yang digunakan di Indonesia adalah Youtube. Banyak artis dan selebgram Indonesia memiliki kanal Youtube untuk mengunggah video aktivitas kesehariannya ataupun jenis konten lainnya seperti video kolaborasi dengan artis lainnya. Tak jarang sesama artis atau selebgram maupun masyarakat umum Indonesia saling melemparkan hujatan terhadap isi konten ataupun pribadi artis tersebut.

Karena tipikal dari *cyberbullying* yang menyerang langsung ke korban maka dampak dari perilaku *cyberbullying* cenderung memiliki luka secara mental baik itu depresi, frustrasi, kehilangan percaya diri [4], dan bahkan tak jarang korban *cyberbullying* berujung pada bunuh diri karena tidak kuat mental atas serangan yang ditujukan pada dirinya [5].

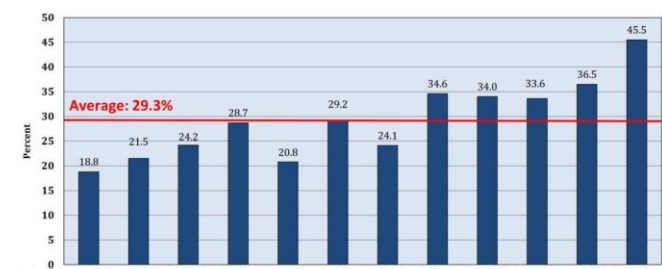
Penegakan hukum tentang kasus *cyberbullying* di Indonesia sebenarnya telah ditetapkan dalam beberapa peraturan seperti [6],[7].

Penelitian tentang deteksi *cyberbullying* telah mulai dilakukan di Indonesia menggunakan variasi metode, bahasa, dan platform media komentar dengan memakai akurasi sebagai tolak ukur pembeding hasil klasifikasi teks [8],[9],[10]. Tetapi penelitian deteksi *cyberbullying* pada media sosial dengan komentar berbahasa Indonesia (serta sebagian bahasa daerah) pada Youtube belum pernah dilakukan.

Penelitian ini bertujuan untuk melakukan deteksi komentar bernada *cyberbullying* pada komentar berbahasa Indonesia (beserta sebagian bahasa daerah) pada media sosial Youtube dengan algoritma *Convolutional Neural Network – Long Short-Term Memory Network (CNN-LSTM)* dengan model utama seperti yang dipakai dalam penelitian sebelumnya [11]. *Dataset* yang dipakai dalam penelitian dikumpulkan dan diberi nama Indonesian Youtube Comment *Dataset (AnOuConDa)*. Rincian *dataset* akan dijelaskan lebih lengkap pada bagian III dari paper ini.

Fitur yang digunakan dalam penelitian ini adalah rangkaian *deep-learning CNN-LSTM* dengan *ontraining word embedding* serta beberapa fitur bantuan untuk meningkatkan akurasi pelabelan.

Paper ini terdiri dari lima bagian. Bagian pertama berisi pendahuluan dan latar belakang penelitian. bagian kedua menjelaskan beberapa penelitian serupa yang telah dilakukan, bagian ketiga dan keempat menjelaskan tentang pengumpulan *dataset* dan proses klasifikasi dari *dataset*, sedangkan bagian kelima menjabarkan hasil serta kesimpulan penelitian.



Gambar 1. Statistik Kasus *Cyberbullying* Dari Tahun 2007-2021

II. PENELITIAN TERKAIT *CYBERBULLYING*

Penelitian pertama yang melakukan klasifikasi *cyberbullying* pada media sosial Twitter dengan 4 macam metode *Supervised Learning* konvensional seperti *Naïve Bayes (NB)*, *Support Vector Machine (SVM)*, *Random Forest*, dan *K-Nearest Network (KNN)* yang dilengkapi dengan pendekatan *synthetic minority oversampling technique (SMOTE)* dan *cost-sensitive* serta membandingkan hasil klasifikasi. Hasil penelitian mengungkapkan bahwa model *random-forest* yang dilengkapi SMOTE memberikan hasil terbaik dengan nilai AUC 0,943 dan *F1-score* 0,936 [12].

Penelitian kedua melakukan klasifikasi *cyberbullying* dengan menggunakan salah satu pendekatan metode *Semi-Supervised Learning* yakni *Fuzzy Support Vector Machine (FSVM)* pada 3 media sosial yakni *Myspace*, *Kongregate*, dan *Slashdot* dan menghasilkan sebuah model FSVM dengan nilai *recall* sebesar 79,3% [13].

Penelitian ketiga menggunakan salah satu metode *Unsupervised Learning* yakni *Growing Hierarchical Self-Organizing Machine (GHSOM)* untuk klasifikasi *cyberbullying* terhadap beberapa *dataset* yang dipakai dalam

penelitian sejenis dan menghasilkan nilai *F1-score* terbaik sebesar 0,74 pada *dataset* Youtube [14].

Penelitian keempat melakukan pendekatan *Deep Learning* dengan menggunakan beberapa *Neural Network* generasi baru seperti *Convolutional Neural Network* (CNN), *Long Short –Term Network* (LSTM), dan *Bidirectional Long Short –Term Network* (BLSTM) dan mengambil kesimpulan bahwa model BLSTM *with attention* memberikan hasil klasifikasi terbaik dengan nilai *F1-score* 0,95 dan akurasi 0,98 [15].

III. INDONESIAN YOUTUBE COMMENT DATASET

Dataset yang digunakan dalam penelitian dibangun dengan mengumpulkan komentar-komentar yang terdapat dalam media sosial Youtube. Pemilihan kanal Youtube hanya dikhususkan kepada orang-orang berkebangsaan Indonesia dan memiliki jumlah komentar kurang lebih sebanyak 1.000 pada video yang akan diambil sebagai *dataset*. Dalam kolom komentar Youtube, sebuah komentar dapat dikomentari kembali oleh diri sendiri maupun pengguna lainnya. Sebanyak 50 video dari tiap kanal orang berkebangsaan Indonesia telah diambil dan dikumpulkan sebanyak total 50.774 data dengan menggunakan API yang telah disediakan dari Youtube sebagai media pengumpulan komentar beserta respon komentar.

Proses pelabelan *dataset* dilakukan oleh 3 orang secara manual dengan hasil akhir 2 label yakni *cyberbullying* dan bukan *cyberbullying*.

Dari *dataset* awal tersebut dilakukan proses pembersihan data kotor (*cleaning*) yakni menghapus data komentar yang mengandung kata berbahasa Inggris dan data kalimat dengan jumlah duplikat diatas satu. *Dataset* hasil proses *cleaning* memiliki ketidakseimbangan jumlah antara label *cyberbullying* dan bukan *cyberbullying* sehingga dilakukan penghapusan secara acak setengah dari data komentar terklasifikasi bukan *cyberbullying* sehingga perbandingan akhir data berlabel *cyberbullying* dengan bukan *cyberbullying* menjadi kurang lebih 1:2.

Dalam proses klasifikasi terdapat beberapa data yang dipakai sebagai pertimbangan apakah suatu komentar tergolong *cyberbullying* yakni data kalimat dan enam fitur tambahan yakni kata-kata umpatan (*swear words*), kata-kata bernada pornografi (*porn related*), kata-kata bernada kasar (*rude words*), kata ganti orang kedua (*second person*), kata-kata bernada *catcalling*, kata-kata bernada melecehkan fisik (*body shaming*), dan terakhir kalimat mengandung *mention* dengan simbol @. Semua fitur tambahan berisi jumlah dari setiap kata yang tergolong dalam tiap fitur. Sebuah kalimat dikatakan sebagai *cyberbullying* apabila minimal 2 dari pelabel menyatakan bahwa kalimat tersebut adalah

Cyberbullying. Setiap kata yang termasuk dalam setiap kategori dicatat dan dikumpulkan kedalam sebuah *vocabulary list* pada saat proses pelabelan termasuk berbagai variasi yang muncul seperti “goblok, goblog, geblek” karena banyaknya kasus *out-of-vocabulary* yang dilakukan oleh orang Indonesia untuk menghindari sensor kata maupun proses *review* komentar oleh pihak YouTube langsung.

Contoh beberapa kata yang tergolong untuk masing-masing fitur dapat dilihat pada Tabel 1.

Dataset akhir berjumlah 26.918 data komentar dengan rincian 17.084 komentar berlabel bukan *cyberbullying* dan 9.834 komentar berlabel *cyberbullying*.

Tabel 1. Contoh Kata Yang Tergolong Setiap Kategori Fitur

Kategori Fitur	Contoh Kata
<i>SwearWords</i>	jancok, asu, anjir, wanjir, njir
<i>PornRelatedWords</i>	bejat, coli, entot, klitoris, kontol
<i>SecondPersonWords</i>	lu, kamu, sampeyan
<i>RudeWords</i>	bacot, belagu, goblog, mampus, taek
<i>CatcallingWords</i>	bening, cantik, gadis, gemoy
<i>BodyShamingWords</i>	gendut, kurus, ketek, paha, sipit

IV. PROSES KLASIFIKASI CYBERBULLYING

A. Praproses

Setelah mendapatkan *dataset* dilakukan beberapa tahapan praproses sebelum data digunakan dalam proses *training* dan *testing*. Praproses dimulai dengan menghilangkan beberapa tanda baca seperti (, . ‘ “ ; : ? !) kecuali tanda @ yang nantinya akan dipakai sebagai salah satu parameter tambahan dalam penelitian. Dilanjutkan dengan menghilangkan semua karakter bergambar / emoji pada setiap komentar. *Stop words* dan angka tidak dihilangkan dalam komentar Terakhir tiap kata dan angka akan diubah menjadi vektor angka yang unik sehingga tidak ada kata dan angka yang mendapatkan vektor yang sama.

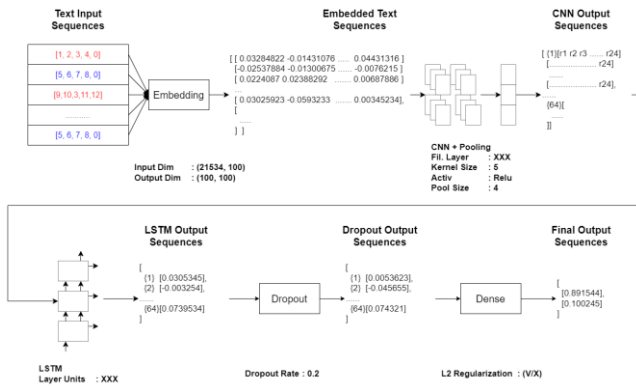
B. Inisiasi Model CNN-LSTM

Model CNN-LSTM yang dipakai menggunakan arsitektur awal yang mirip seperti penelitian mengenai *cyberbullying* sebelumnya. Model ini dipilih karena memberikan hasil lebih baik secara *F1-score* dibandingkan beberapa model lainnya seperti *Support Vector Machine* (SVM), *Logistic Regression* (LR), dan bahkan model kompleks sekalipun yakni gabungan dari CNN-LSTM-DNN meski memakai variasi *dataset* yang tidak seimbang jumlah klasifikasinya maupun yang mendekati seimbang.

Model diawali dengan layer *on-training embedding* yang menghasilkan matriks sebesar (100,100) kemudian dilanjutkan dengan CNN dengan jumlah layer tertentu dengan penambahan *maxPooling* 1 dimensi dengan besar *pool* 4, ukuran *kernel* sebanyak 5 dan diikuti oleh layer LSTM yang jumlah *unit layer*-nya sama seperti CNN yang akan dibuat bervariasi dalam penelitian. Sebuah *dropout layer* menerima hasil matriks dari LSTM dan melakukan *dropout* dengan probabilitas 0,2 dan disambungkan dengan sebuah *dense layer* yang bertugas sebagai *output layer* yang memberikan nilai probabilitas masing-masing label untuk setiap hasil klasifikasi. Model menggunakan *learning rate* sebesar 0,001 dan algoritma Adam sebagai optimasi dalam model CNN-LSTM.

Penelitian ini menggunakan 2 macam variasi model yakni variasi jumlah *filter layer* CNN sekaligus *layer unit* LSTM dan ada tidaknya *L2 Regularization* pada *output layer* untuk mendapatkan model terbaik untuk klasifikasi komentar *cyberbullying*. Nilai untuk variasi jumlah *layer* CNN dan LSTM didapatkan dari pola pangkat dari 2 (2^x) dimana beberapa nilai yang dipakai dalam penelitian yakni 128, 256, 512, dan 1024. Gambaran arsitektur model CNN-LSTM yang dipakai dalam penelitian dapat dilihat pada Gambar 2. Dalam penelitian ini digunakan tipe *on-training embedding* dikarenakan beberapa hal berikut:

- Data komentar pengguna Youtube cenderung memiliki beberapa variasi terhadap kata yang tergolong sebagai *cyberbullying* seperti “bangsat”, “bangsad”, “crot”, “cрут”, dan banyak variasi lainnya lagi yang menyebabkan kasus OOV umum terjadi. *On-training embedding* yang selalu diperbarui dalam tiap iterasi memungkinkan *embedding layer* untuk lebih adaptif dan ikut mempelajari kata-kata serupa dalam *dataset*.
- Variasi bahasa konten *cyberbullying* serta makna kata yang dapat berubah sewaktu-waktu menjadi tantangan tersendiri untuk *embedding layer* agar dapat beradaptasi dan lebih relevan serta akurat dalam model. Disinilah *on-training embedding* menjadi solusi untuk menyelesaikan masalah tersebut.



Gambar 2. Arsitektur Model CNN-LSTM

C. Penanganan Out-Of-Vocabulary Pada Dataset

Penanganan *out-of-vocabulary* (OOV) menggunakan OOV *token* dari *library* Keras dengan mengganti semua kata yang tidak dikenal menjadi nilai “<OOV>”. Penggunaan metode ini juga berintegrasi dengan penggunaan *on-training embedding* dimana model dapat lebih adaptif untuk menangani OOV pada data yang belum pernah dilihat sebelumnya.

D. Proses Training Dan Testing

Proses *training* dimulai dengan meng-*impor training dataset* berupa file berekstensi .csv serta mendeklarasi semua *library* yang digunakan yakni *tensorflow*, *keras*, *sklearn*, *nlTK* dan *library* standar *python* sebagai media proses model. Setelah selesai melakukan inialisasi parameter setiap data akan mulai masuk kedalam model CNN-LSTM. Model akan melakukan proses *training* pada *dataset* yang telah dipisah

secara acak oleh bantuan fungsi *train_test_split* dengan perbandingan data *training* sebanyak 80 persen dan data *testing* sebanyak 20 persen. Proses *training* dilakukan selama 10 *epoch* dan dilakukan sebanyak 5 kali dengan menggunakan metode *k-fold cross validation* sebagai bentuk validasi untuk model tersebut.

Proses *training* dilakukan ke semua variasi model baik jumlah *layer* baik model CNN dan LSTM dan ada tidaknya *L2 Regularization* dan mendapatkan hasil perbandingan nilai akurasi rata-rata, akurasi *final*, dan *loss final* seperti tertera pada Tabel 2. Model dengan variasi jumlah *layer* 1024 memberikan nilai akurasi rata-rata tertinggi meski memberikan nilai akurasi *final* yang lebih rendah bila dibandingkan dengan model dengan variasi jumlah *layer* 128. Penambahan *L2 Regularization* pada model tidak memberikan peningkatan performa tetapi justru menurunkan seluruh nilai tolak ukur bila dibandingkan dengan model tanpa *L2 Regularization*.

Tabel 2. Perbandingan Berbagai Nilai Tolak Ukur Hasil Training Tiap Variasi Model

Model	Akurasi Avg.	Akurasi Final	Loss Final
256 Filter Layer	0,9654	0,9958	0,0169
256 Filter Layer + L2 Regularization	0,9581	0,9953	0,0268
512 Filter Layer	0,9652	0,9949	0,0177
1024 Filter Layer	0,9658	0,9965	0,0174
128 Filter Layer	0,9650	0,9970	0,0090

Setelah model selesai melakukan *training* maka sekarang giliran data *testing* yang dimasukkan untuk diuji performa untuk setiap model. Hasil dari proses klasifikasi pada data *testing* dicatat dalam *confusion matrix* untuk mendapatkan nilai tolak ukur *precision*, *recall*, dan *F1-score* dimana salah satu contohnya terlihat pada Tabel 3.

Tabel 3. Salah satu Confusion Matrix dari Model variasi 1024 Layer

	Predicted Negative	Predicted Positive
Actual Negative	3242	190
Actual Positive	397	1555

Pada bagian kolom data *cyberbullying* yang benar terdeteksi sebagai *cyberbullying* (*true positive*) berjumlah sekitar setengah dari bukan *cyberbullying* dikarenakan ketidakseimbangan jumlah *dataset* sebelum proses pembagian data *training* dan *testing* dilakukan.

Hasil percobaan data *testing* pada setiap variasi model dapat dilihat pada Tabel 4 dimana model dengan variasi jumlah *layer* sebanyak 1024 menghasilkan nilai *F1-score* terbaik sebesar 0,841. Sedangkan pada model berjumlah *layer* 256 yang ditambahkan *L2 Regularization* justru mengalami penurunan nilai *F1-score* pada data *testing*. Penambahan jumlah *layer* untuk kedua model memberikan peningkatan nilai *F1-score* dalam jumlah kecil (sekitar 0,005)

Tabel 4. Perbandingan Berbagai Nilai Tolak Ukur Hasil Pengujian Dengan Data *Testing*

	Akurasi Avg.	Precision	Recall	F1-Score
256 Filter Layer	0,9654	0,850	0,813	0,831
256 Filter Layer + L2 Regularization	0,9581	0,843	0,817	0,830
512 Filter Layer	0,9652	0,893	0,785	0,836
1024 Filter Layer	0,9658	0,891	0,797	0,841
128 Filter Layer	0,9650	0,859	0,813	0,835

Model yang telah selesai dibuat kemudian dipakai dalam sebuah *Application Programming Interface* (API) sederhana yang dibuat dengan menggunakan *library flask* agar device pengguna dapat menerima request API via *localhost* dan diuji dengan menggunakan aplikasi *Postman* yang menerima *input* dalam bagian *body* dari *POST request* berupa beberapa kalimat untuk dapat dicoba dideteksi konten *cyberbullying*-nya (Segmen Program 1 baris 1 hingga 6) dengan ekstensi API */sentiment* dan menghasilkan *output* berupa *list* hasil deteksi untuk setiap data kalimat yang dikirimkan (Segmen Program 1 baris 7 hingga 13).

```

01:      {
02:          "comments_1": "konten sampah modal
      bacot cok",
03:          "comments_2": "pamer tetek sama
      dada",
04:          "comments_3": "kenikmatan tiada
      tara"
05:      }
06:
07:      {
08:          "comments": {
09:              "comments_1": "Cyberbullying",
10:              "comments_2": "Cyberbullying",
11:              "comments_3": "Bukan Cyberbullying"
12:          }
13:      }
    
```

Segmen Program 1. Contoh *Input* dan *Output* API *Postman* Dengan Menggunakan Model CNN-LSTM

V. KESIMPULAN

Klasifikasi komentar *cyberbullying* dengan menggunakan Model CNN-LSTM memberikan hasil deteksi terbaik dengan nilai *F1-score* 0,841. Model tersebut dapat digunakan dalam proses deteksi menggunakan API yang menerima beberapa *input* kalimat yang akan dideteksi konten *cyberbullying*-nya. Kendala yang sering dihadapi dalam penelitian *cyberbullying* adalah ketidakseimbangan jumlah data *cyberbullying* yang cenderung lebih sedikit dibandingkan dengan data bukan *cyberbullying*. Hal ini dapat diatasi dengan pengurangan jumlah data bukan *cyberbullying* agar model tidak terlalu bias dalam proses deteksi.

Penelitian tentang klasifikasi *cyberbullying* di Indonesia pada berbagai media sosial memiliki banyak potensi untuk pengembangan mengingat masyarakat Indonesia sering

berselancar di berbagai media sosial dan tak jarang meninggalkan komentar.

REFERENSI

- [1] Rahayu, F. S. (2012). Cyberbullying sebagai dampak negatif penggunaan teknologi informasi. *Journal of Information Systems*, 8(1), 22-31.
- [2] C. Salmivalli, "Bullying and the Peer Group: A Review," *Aggression and Violent Behavior*, vol. 15, no. 2, pp. 112–120, 2010. doi:10.1016/j.avb.2009.08.007.
- [3] [1] J. W. Patchin, "Summary of our cyberbullying research (2004-2022)," Cyberbullying Research Center, <https://cyberbullying.org/summary-of-our-cyberbullying-research>.
- [4] Hinduja, S. & Patchin, J. W. (2020). *Cyberbullying fact sheet: Identification, Prevention, and Response*. Cyberbullying Research Center. Retrieved May 15, 2020, from <https://cyberbullying.org/Cyberbullying-Identification-Prevention-Response-2020.pdf>
- [5] Dama, A. (2019, October 20). *Artis Korea Sulli Bunuh Diri Karena Dibully, Inilah Isi Deretan Komentar Bully Sebut Pelacur*. Retrieved May 15, 2020, from <https://kupang.tribunnews.com/2019/10/20/artis-korea-sulli-bunuh-diri-karena-dibully-inilah-isi-deretan-komentar-bully-sebut-pelacur>
- [6] N. R. Muhammad and S. Nandang, "Perlindungan Hukum terhadap Korban Cyberbullying pada Remaja di Tinjau dari Undang-Undang Nomor 19 Tahun 2016 tentang Perubahan atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik," *Prosiding Ilmu Hukum*, vol. 7, no. 2, pp. 775–780, Aug. 2021. doi:10.29313/v0i0.27613
- [7] A. Sakban, S. Sahrul, A. Kasmawati, and H. Tahir, "Kebijakan Hukum Pidana Terhadap Kejahatan cyber bullying di Indonesia," *CIVICUS: Pendidikan-Penelitian-Pengabdian Pendidikan Pancasila dan Kewarganegaraan*, vol. 7, no. 2, p. 59, 2019. doi:10.31764/civicus.v7i2.1195
- [8] L. Anindyati, A. Purwarianti, and A. Nursanti, "Optimizing Deep Learning for detection cyberbullying text in Indonesian language," *2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, 2019. doi:10.1109/icaicta.2019.8904108
- [9] J. Pardede, "Deteksi Komentar cyberbullying Pada media sosial Berbahasa Inggris menggunakan naïve bayes classification," *Jurnal Informatika*, vol. 7, no. 1, pp. 46–54, 2020. doi:10.31311/ji.v7i1.6920
- [10] R. Masbadi Hatullah Nurnaryo et al., "Deteksi Cyberbullying Pada data tweet Menggunakan metode random forest Dan Seleksi FITUR Information Gain," *Jurnal Simantec*, vol. 11, no. 1, pp. 33–40, 2022. doi:10.21107/simantec.v11i1.17256
- [11] H. Rosa, D. Matos, R. Ribeiro, L. Coheur, and J. P. Carvalho, "A 'deeper' look at detecting cyberbullying in social networks," *2018 International Joint*

-
- Conference on Neural Networks (IJCNN)*, 2018. doi:10.1109/ijcnn.2018.8489211
- [12] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network," *Computers in Human Behavior*, vol. 63, pp. 433–443, 2016. doi:10.1016/j.chb.2016.05.051
- [13] V. Nahar, S. Al-Maskari, X. Li, and C. Pang, "Semi-supervised Learning for cyberbullying detection in social networks," *Lecture Notes in Computer Science*, pp. 160–171, 2014. doi:10.1007/978-3-319-08608-8_14
- [14] M. Dittenbach, D. Merkl and A. Rauber, "The growing hierarchical self-organizing map," *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, Como, Italy, 2000, pp. 15-19 vol.6, doi: 10.1109/IJCNN.2000.859366.
- [15] S. Agrawal and A. Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms," *Lecture Notes in Computer Science*, pp. 141–153, 2018. doi:10.1007/978-3-319-76941-7_11