

Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM

by Habib Faizal Fadli

Submission date: 24-Nov-2020 04:55PM (UTC+0700)

Submission ID: 1454965957

File name: PAPER_TANPA_NAMA_PENULIS.pdf (627.52K)

Word count: 2742

Character count: 17468

Identifikasi *Cyberbullying* pada Media Sosial *Twitter* Menggunakan Metode LSTM dan BiLSTM

Abstract—*Cyberbullying* merupakan masalah yang harus menjadi perhatian penting oleh masyarakat. *Cyberbullying* termasuk kebiasaan buruk yang berdampak mengerikan, mulai dari gangguan psikologis korban, hingga munculnya kasus bunuh diri. Tujuan dari penelitian ini adalah mengidentifikasi konten yang mengandung makna perundungan secara daring (*cyberbullying*) pada media sosial khususnya *Twitter*. Dalam kasus ini, sumber data penelitian ini berasal dari media sosial *Twitter*. Setidaknya, ada 6835 data yang telah dikumpulkan. Data tersebut terdiri dari dua jenis cuitan dengan masing-masing cuitan memiliki kecenderungan *cyberbullying* dan *non-cyberbullying*. Tujuan penelitian akan tercapai dengan melakukan beberapa langkah yang pertama yaitu pengumpulan data, lalu *preprocessing*, kemudian ekstraksi fitur, setelah itu klasifikasi, dan terakhir evaluasi. Dua algoritma *deep learning* diimplementasikan dalam penelitian ini, yaitu LSTM dan BiLSTM. Sehingga, disimpulkan bahwa kedua algoritma tersebut memiliki performa yang relatif sama. Akurasi dari masing-masing algoritma dituliskan sebagai berikut *Long Short Term Memory* 81,60 % dan *Bidirectional Long Short Term Memory* 81,78 %. Lalu, untuk nilai dari *F1-Score* dari masing-masing algoritma sebagai berikut *Long Short Term Memory* 77,88 % dan *Bidirectional Long Short Term Memory* 77,89 %.

Keywords—*Twitter*, *Cyberbullying*, *Deep Learning*, *Klasifikasi*

I. PENDAHULUAN

Media sosial merupakan tempat dimana penggunanya melakukan interaksi dengan pengguna lain secara daring tanpa mengenal waktu dan tempat. Indonesia sendiri merupakan negara dengan jumlah pengguna media sosial tertinggi di dunia. Berdasarkan Kementerian Komunikasi dan Informatika (Kemenkominfo) menyatakan 95% dari sekitar 63 juta pengguna internet adalah pengguna media sosial [1]. *Twitter* merupakan aplikasi yang sering digunakan oleh masyarakat Indonesia. Menurut *Country Industry Head Twitter Indonesia* mengklaim bahwa Indonesia termasuk negara dengan pertumbuhan pengguna aktif harian *Twitter*-nya paling besar [2].

Penggunaan *Twitter* di Indonesia tidak hanya ditujukan untuk perorangan. Akan tetapi, *Twitter* juga digunakan oleh lembaga-lembaga negara, komunitas, hingga toko *online*. Namun, tidak semua pengguna menggunakan teknologi ini dengan bijak. Tindakan negatif seperti penipuan, penyebaran *hoax*, menyebarkan opini yang cenderung mengandung ujaran kebencian, hingga perundungan secara daring (*cyberbullying*) banyak dilakukan oleh pengguna *twitter* [3]. Selama tahun 2019 masyarakat telah melaporkan setidaknya 244.738 jumlah konten pornografi. Lalu, terdapat sejumlah 19.970 konten kategori perjudian. Ada pula konten penipuan sejumlah 18.845, dan konten informasi hoaks, serta konten mengandung SARA, terorisme, radikalisme, pelanggaran HAK, serta kekerasan terhadap anak sejumlah 15.361 [4]. Hal ini tentu menjadi dampak negatif penggunaan media sosial. Dampak-dampak yang dapat ditimbulkan seperti kerusakan karena menyebarnya berita bohong, kerugian materiil karena penipuan, hingga berbagai kasus yang ditimbulkan karena *cyberbullying* [5].

Cyberbullying atau perundungan secara daring adalah tindakan penyerangan, penghinaan, atau menyakiti orang lain secara sengaja dan berulang-ulang pada sosial media, pesan, atau dengan cara lainnya [5]. *Cyberbullying* di media sosial *Twitter* dilakukan dengan menulis cuitan yang mengandung kata-kata hinaan atau kata-kata kasar, bahkan kata-kata yang menjerus kepada penghinaan terhadap SARA. Pemerintah mengumumkan setidaknya 84% remaja berusia 12 sampai 17 tahun di Indonesia menjadi korban tindakan perundungan (*bullying*) dan kebanyakan kasus *bullying* yang ditemukan merupakan *cyberbullying*[6]. *Cyberbullying* menjadi kekhawatiran publik karena banyak kasus *cyberbullying* sering dikaitkan dengan tindakan bunuh diri. Salah satu organisasi non-profit, *Cyber Bullying Research Center* mengungkapkan bahwa kebiasaan *bullying* baik secara langsung maupun secara daring di kalangan remaja dapat mengakibatkan depresi, tindakan bunuh diri dan percobaan pembunuhan [5]. Oleh karena efek berbahaya yang ditimbulkan *cyberbullying*, tindakan pencegahan atau deteksi perlu dilakukan agar tidak membahayakan korban maupun pelaku.

Penelitian dilakukan dengan melakukan lima tahapan proses. Langkah pertama yaitu pengumpulan data, lalu *preprocessing*, kemudian ekstraksi fitur, setelah itu klasifikasi, dan terakhir evaluasi. Metode yang digunakan yaitu *deep learning* dan algoritma yang digunakan yaitu LSTM dan BiLSTM. Penelitian ini penting dilakukan karena belum ditemukan model yang dibuat khusus untuk mendeteksi dan mengklasifikasikan cuitan yang berbahasa Indonesia dan bermakna *cyberbullying* dengan metode *deep learning* terutama dengan 2 algoritma tersebut. Kedua algoritma tersebut dinilai sangat tepat dalam melakukan deteksi dan klasifikasi dibandingkan algoritma-algoritma yang lainnya.

Penelitian ini mengklasifikasikan data untuk mendeteksi *cyberbullying* di media sosial *Twitter* menjadi dua kelas yaitu *cyberbullying* dan bukan *cyberbullying*. Kelas *cyberbullying* berisi cuitan berupa kata-kata yang mengandung unsur *cyberbullying*, sementara kelas bukan *cyberbullying* berisi cuitan berupa kata-kata yang tidak mengandung unsur *cyberbullying*. Dengan dikembangkannya penelitian ini, diharapkan akan dapat membantu orang tua, pemerintah, dan negara untuk melindungi generasi muda dari perundungan secara daring (*cyberbullying*) dan menekan jumlah para pelaku *cyberbullying*.

II. TINJAUAN PUSTAKA

Penelitian mengenai klasifikasi teks sudah banyak dilakukan oleh peneliti-peneliti sebelumnya. Konten kasar yang diklasifikasikan berdasarkan cuitan media sosial *Twitter* dengan kamus bahasa Indonesia juga pernah dilakukan sebelumnya. Hidayatullah et al. melakukan klasifikasi berdasarkan cuitan menjadi dua kelas dan melakukan perbandingan terhadap performa algoritma NBC dan SVM dalam melakukan klasifikasi [7]. Berikut adalah hasil *accuracy*, *Precision*, *Recall*, dan *F1-Score* dari masing-masing model yang digunakan dimana NBC mempunyai nilai 0.9834; 0.9912; 0.9762; 0.9836 dan SVM mempunyai nilai 0.9928; 0.9914; 0.9946; 0.9930.

Penelitian tentang deteksi *cyberbullying* pada media sosial *Twitter* berbahasa Indonesia pernah dilakukan sebelumnya. Abdullah et al. melakukan deteksi cuitan pada media sosial *Twitter*. Data yang berupa cuitan tersebut nantinya akan dimasukkan ke dalam kelas *cyberbullying* dan kelas *non-cyberbullying*. Deteksi cuitan tersebut menggunakan metode *Machine Learning* dengan algoritma NBC, *SVM*, *Logistic Regression*, dan *KNN* [8]. Berikut adalah hasil *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dari masing-masing algoritma yang digunakan dimana algoritma NBC mempunyai nilai 0.961; 0.96; 0.96; 0.96, *SVM* mempunyai nilai 0.994; 0.99; 0.99; 0.99, *Logistic Regression* mempunyai nilai 0.997; 1.00; 1.00; 1.00, dan *KNN* mempunyai nilai 0.918; 0.93; 0.92; 0.92.

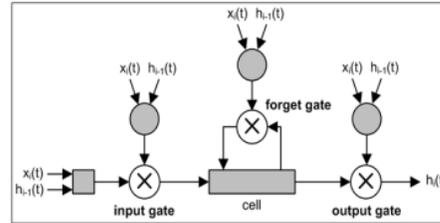
Penelitian tentang deteksi *cyberbullying* pada media sosial selain *Twitter* juga sudah pernah dilakukan sebelumnya. Homa Hosseinmardi et al. melakukan penelitian tentang klasifikasi *cyberbullying* menggunakan algoritma NBC dan *SVM* pada media sosial Instagram berdasarkan foto dan komentar [9]. Berikut adalah hasil *Accuracy*, *Precision*, dan *Recall* dari masing-masing algoritma yang digunakan dimana algoritma *SVM* mempunyai nilai 0.74; 0.74; 0.78 dan *SVM* mempunyai nilai 0.87; 0.88; 0.87.

Ada pula penelitian tentang deteksi *cyberbullying* yang digabungkan dengan metode pendekatan psikologi yang pernah dilakukan sebelumnya. Balakrishnan et al. melakukan penelitian ini menggunakan model *Big Five* and *Triad* dan algoritma *Random Forest* [10]. Berikut adalah hasil *Precision*, *Recall*, dan *F-Measure* dari algoritma yang digunakan dimana algoritma *Random Forest* mempunyai nilai 0.960; 0.952; 0.929.

Penelitian tentang *bullying* yang mengambil data lebih dari satu media sosial juga pernah dilakukan sebelumnya. Agrawal et al. melakukan penelitian tentang *cyberbullying* di tiga media sosial yaitu *Formspring*, *Twitter*, dan *Wikipedia* dengan membagi 4 kelas *bullying* yaitu *bully*, *racism*, *sexism*, dan *attack* kemudian menguji data dengan mengklasifikasinya dengan algoritma *CNN*, *LSTM*, *BiLSTM*, dan *BiLSTM with attention*. Akan tetapi, pada penelitian tersebut masih menggunakan korpus bahasa Inggris [11]. Berikut adalah hasil *Precision*, *Recall*, dan *F1-Score* dari masing-masing algoritma yang digunakan dimana *CNN* mempunyai nilai 0.93; 0.90; 0.91, *LSTM* mempunyai nilai 0.91; 0.85; 0.88, *BiLSTM* mempunyai nilai 0.91; 0.81; 0.86, dan *BiLSTM with attention* mempunyai nilai 0.90; 0.91; 0.91.

III. LONG SHORT TERM MEMORY

LSTM adalah modifikasi dari *Recurrent Neural Network* (*RNN*) dengan adanya penambahan *memory cell* yang digunakan untuk menyimpan informasi dengan jangka waktu yang panjang, serta *LSTM* juga dapat menangani masalah *vanishing gradient* yang terdapat pada *RNN* saat memproses data sekuensial yang panjang dengan menggunakan satu set gerbang yang digunakan untuk mengontrol informasi yang masuk ke memori [12].

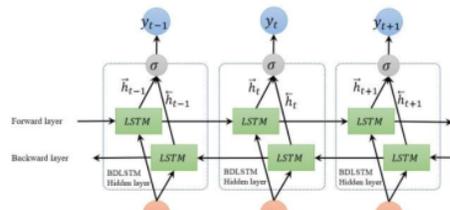


Gambar 1. Arsitektur LSTM

Pada Gambar 1 menunjukkan arsitektur dari *BiLSTM*. *Cell state* merupakan tempat untuk menyimpan informasi yang diberikan dari satu langkah waktu ke langkah waktu berikutnya. *Gate units* berperan dalam memproses informasi yang dibutuhkan dan dibuang. *Gate units* terdiri dari *input gate*, *forget gate*, dan *output gate*. *Input gate* merupakan *gate* yang berfungsi untuk memutuskan nilai *input* yang akan diteruskan pada *cell state* untuk diperbaharui. *Forget gate* merupakan *gate* yang memutuskan informasi mana yang perlu dibuang dari *cell state*. *Output gate* merupakan *gate* yang memutuskan *output* yang akan dihasilkan. Pada Gambar 1 menunjukkan alur *gate units* dalam mengontrol konektivitas *LSTM*.

IV. BIDIRECTIONAL LONG SHORT TERM MEMORY

BiLSTM adalah perkembangan dari model *LSTM* dimana terdapat dua lapisan yang prosesnya saling berkebalikan arah, model ini sangat baik untuk mengenali pola dalam kalimat karena setiap kata dalam dokumen diproses secara sekuensial, karena cuitan dapat dipahami bila pembelajaran secara berurut setiap kata. Lapisan dibawahnya bergerak maju (*forward*), yaitu memahami dan memproses dari kata pertama menuju kata terakhir sedangkan lapisan diatasnya bergerak mundur (*backward*), yaitu memahami dan memproses dari kata terakhir menuju kata pertama. Dengan adanya lapisan dua arah yang saling berlawanan ini maka model dapat memahami dan mengambil perspektif dari kata terdahulu dan kata terdepan, sehingga proses pembelajaran akan semakin dalam yang berdampak pada model akan lebih memahami konteks pada cuitan tersebut.



Gambar 2. Arsitektur BiLSTM

Pada Gambar 2 dapat dilihat bahwa setiap *hidden unit* keluaran *unit* pada lapisan bawah dan atas di digabungkan membentuk nilai fitur kata tersebut dengan ukuran lebih panjang daripada menggunakan *LSTM* biasa. Karena lebih panjang nilai fitur, maka informasi yang akan di proses pada tahap selanjutnya yaitu *feed forward neural* akan mengklasifikasikan dengan lebih akurat.

BiLSTM akan sangat bermanfaat dalam hal pelabelan sekuensial apabila memiliki akses terhadap kedua informasi dari sebelum dan sesudahnya. Namun *hidden state* pada LSTM hanya mengambil informasi dari sebelumnya (masa lalu), sedangkan untuk informasi yang ada setelahnya tidak diketahui. Permasalahan tersebut dapat dipecahkan dengan menggunakan BiLSTM [13]. Pada dasarnya BiLSTM terdiri dari dua LSTM, *forward LSTM* dan *backward LSTM*, sehingga gabungan tersebut akan menangkap informasi dari kedua arah.

V. METODOLOGI

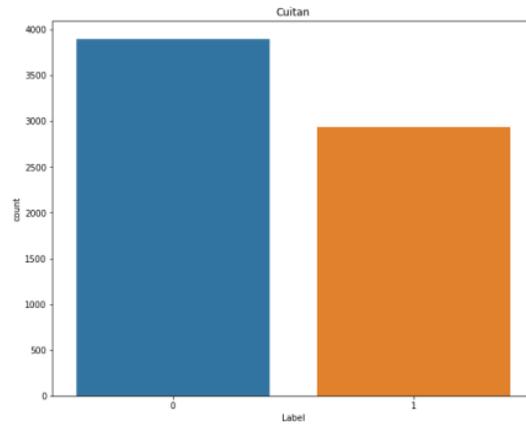
4 Pada bagian ini berisi pokok bahasan yang menjelaskan tentang langkah-langkah yang digunakan dalam penelitian ini. Pertama yaitu pengumpulan data, lalu *preprocessing*, kemudian ekstraksi fitur, setelah itu klasifikasi, dan terakhir evaluasi. Pada Gambar 3 seperti yang terlihat akan dijelaskan pada sub bab berikutnya.



Gambar 3. Alur Penelitian

A. Pengumpulan Data

Twitter adalah tempat pengambilan data untuk penelitian ini. Langkah pengumpulan data dilakukan dengan memanfaatkan Twitter API. Sebelum menggunakan Twitter API, penulis harus mendaftar mendaftarkan diri sebagai *developer* untuk mendapatkan izin akses berupa *Consumer Key*, *Consumer Secret*, *Access Token*, dan *Access Secret*. Data yang akan diambil dari Twitter berupa data cuitan berbahasa Indonesia. Cuitan yang terkumpul untuk membangun model pada penelitian ini adalah 6835 cuitan seperti yang terlihat pada Gambar 4 dengan cacah masing masing yaitu 3900 cuitan *non-cyberbullying* dan 2935 cuitan *cyberbullying*.



Gambar 4. Cacah Data

Cuitan *non-cyberbullying* diambil menggunakan kata kunci yang positif. Sedangkan cuitan *cyberbullying* diambil dengan menggunakan kata kunci yang merupakan kata - kata yang bermakna merundungan.

B. Preprocessing

Preprocessing merupakan langkah dimana membersihkan data mentah berupa cuitan-cuitan sehingga menjadi data yang baik dan terstruktur. Langkah *preprocessing* yang digunakan merujuk pada langkah-langkah *preprocessing* yang dilakukan Hidayatullah et al [14]. Berikut langkah - langkah *preprocessing* yang dilakukan antara lain:

- Menghilangkan *link* atau URL.
- Menghilangkan karakter NON-ASCII.
- Menghilangkan angka, simbol dan tanda baca.
- Menghilangkan *hashtag*, *username*, dan *retweet*.
- Menghilangkan huruf ke dalam bentuk *lowercase*.
- Menghilangkan *stopwords*.
- Merubah Bahasa informal ke Bahasa formal.
- Menghilangkan kata yang terdiri dari satu huruf.
- Menghilangkan digit.

10 Berikut merupakan hasil dari tahap *preprocessing* cuitan yang dapat dilihat pada tabel 1.

Tabel 1. *Preprocessing*

Sebelum	Sesudah
@pengguna 1 @pengguna 2 dihh itu artis sok ngartis ngartis	itu artis sok ngartis
@pengguna 3 heloow yg sombong gaya gapernah ngerasain ditampol spokat	yang sombong gaya gapernah ngerasain ditampol spokat

C. Ekstraksi Fitur

Ekstraksi fitur adalah proses untuk menggambarkan karakteristik dari sebuah objek [15]. Ekstraksi fitur bertujuan

untuk menggali informasi yang berguna dari sampel asli dan merepresentasikannya sebagai vektor fitur yang dinormalisasi dengan ukuran yang sama. Metode ekstraksi fitur yang efektif biasanya membantu meningkatkan kinerja sistem prediksi [16]. Fitur juga dapat membantu dalam menentukan sebuah data masuk ke kategori kelas yang mana [17]. Ekstraksi fitur dilakukan menggunakan metode *Word2Vec*.

Word2Vec akan mengubah kata-kata ke dalam bentuk vektor. Terdapat dua algoritma pembelajaran yang utama pada *Word2Vec*, yaitu CBOW dan *Continuous Skip-gram*. Pada arsitektur CBOW, model akan memprediksi kata saat ini berdasarkan konteks data di sekitarnya. Sedangkan pada arsitektur *Skip-gram*, model menggunakan kata saat ini untuk memprediksi konteks kata di sekitarnya. Tujuan pelatihan model CBOW adalah untuk menggabungkan representasi kata di sekitarnya untuk memprediksi kata yang ada di tengah, sedangkan tujuan pelatihan model *Skip-gram* adalah untuk mempelajari representasi vektor kata yang baik pada saat memprediksi konteksnya dalam kalimat yang sama, dan karena memiliki kompleksitas komputasi yang rendah, kedua model tersebut dapat dilatih pada korpus yang berukuran besar dalam waktu singkat [18].

D. Klasifikasi

Langkah ini akan menjelaskan tentang proses klasifikasi cuitan dengan menggunakan beberapa algoritma *deep learning* yaitu LSTM dan BiLSTM. Data yang berupa teks diklasifikasikan ke dalam kelas-kelas *cyberbullying* menggunakan algoritma tersebut karena algoritma tersebut terbukti karena memiliki akurasi yang sangat baik dalam klasifikasi [11]. Klasifikasi pada penelitian ini data cuitan akan dibagi menurut 2 kelas yaitu kelas *cyberbullying* dan kelas *non-cyberbullying*. Data yang berupa cuitan nantinya akan diklasifikasikan menggunakan algoritma yang sudah ditentukan yaitu LSTM dan BiLSTM.

E. Evaluasi

Confusion matrix merupakan matriks yang menyimpan informasi untuk mengetahui performa dari model yang digunakan dan digunakan sebagai acuan dari performa klasifikasi dari algoritma yang digunakan pada tahap evaluasi [7].

Tabel 2. *Confusion Matrix*

		Predicted Values	
		Positive (0)	Negative (1)
Actual Values	Positive (0)	TP	FP
	Negative (1)	FN	TN

Confusion Matrix adalah sumber informasi apakah model yang digunakan berkinerja baik atau tidak. Hal ini bisa dilihat dari nilai yang dari variabel TP (*True Positive*) dan variabel TN (*True Negative*) menunjukkan total prediksi benar yang dibuat oleh model. Sedangkan nilai pada variabel FP (*False Positive*) dan variabel FN (*False Negative*) menunjukkan total prediksi salah yang dibuat oleh model. Penghitungan kinerja model dapat dilakukan dengan menghitung nilai *accuracy*,

precision, *recall*, dan *F1-Score* berdasarkan rumus yang terlihat pada persamaan (1), (2), (3) dan (4).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

keterangan :

TP = Jumlah data kelas positif (0) diprediksi benar sebagai kelas positif (0)

FN = Jumlah data kelas positif (0) diprediksi salah sebagai kelas negatif (1)

TN = Jumlah data kelas negatif (1) diprediksi benar sebagai kelas negatif (1)

FP = Jumlah data kelas negatif (1) diprediksi salah sebagai kelas positif (0)

VI. HASIL DAN PEMBAHASAN

A. Skenario

Semua model akan menggunakan *embedding layer* dengan ukuran 64 dimensi. Semua layer dengan 64 *units* dan sebuah *dense layer* dimana fungsi aktivasinya menggunakan *Softmax* dan fungsi optimasinya menggunakan *Adam*. Semua model akan dijalankan menggunakan 20 *epoch* dan 32 *batch size*.

Hasil dari model akan dibandingkan satu sama lain untuk melihat model manakah yang terbaik dalam klasifikasi dengan melihat beberapa parameter yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Berdasarkan hasil cacah *dataset* menunjukkan ketidakseimbangan yaitu 43% banding 57% sehingga perlukan beberapa parameter di atas untuk melihat model manakah yang memiliki performa terbaik.

B. Performa Model

Data yang sudah dibersihkan pada langkah *preprocessing* selanjutnya dibagi menjadi dua yaitu data *testing* dan data *training* dengan cacah masing-masing 0.2 data *testing* dan 0.8 data *training*. Berikut disajikan tabel hasil dari masing-masing *confusion matrix* algoritma *deep learning*.

Tabel 3. *Confusion Matrix LSTM*

Long Short Term Memory (LSTM) Network			
		Predicted Value	
		0	1
Actual Values	0	2152	337
	1	468	1417

Tabel 4. *Confusion Matrix BiLSTM*

Bidirectional Long Short Term Memory (BiLSTM) Network	
Predicted Value	

<i>Actual Values</i>		0	1
	0	2173	316
	1	481	1404

Berdasarkan *confusion matrix* yang sudah didapat selanjutnya nilai dari *accuracy*, *precision*, *recall*, dan *F1-Score* dapat diketahui. Berikut merupakan nilai dari masing-masing variabel tersebut pada tabel.

Tabel 5. *Accuracy*, *Precision*, *Recall*, dan *F1-Score*

Metode	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
LSTM	81,60	80,79	75,17	77,88
BiLSTM	81,78	81,63	74,48	77,89

Berdasarkan tabel 5, algoritma BiLSTM terbukti lebih unggul dibanding dengan LSTM baik dari segi dari *accuracy*, *precision*, *recall*, maupun *F1-Score*.

VII. KESIMPULAN

Berdasarkan hasil yang telah dicapai, penelitian ini telah berhasil melakukan identifikasi cuitan bermakna

cyberbullying pada media sosial *Twitter* dengan melakukan klasifikasi antara dua kelas cuitan yang tersedia pada *dataset*.

Hasil evaluasi dari masing - masing algoritma *deep learning* menempatkan *Bidirectional Long Short Term Memory* sebagai algoritma terbaik dalam mengklasifikasi data cuitan dibanding dengan *Long Short Term Memory*. Hal ini dibuktikan dengan nilai dari *accuracy*, *precision*, *recall*, dan *F1-Score* dari BiLSTM yaitu 81.78; 81.63; 74.48; 77.89

Namun, *Long Short Term Memory* juga algoritma yang cukup baik dalam mengklasifikasi data teks. Ini dibuktikan dengan nilai dari *accuracy*, *precision*, *recall*, dan *F1-Score* yang selisih sedikit nilainya dengan *Bidirectional Long Short Term Memory*.

Penelitian ini masih sebatas model dalam klasifikasi cuitan yang bermakna *cyberbullying* pada media sosial. Penulis berharap penelitian ini bisa dikembangkan dengan melakukan menambah *dataset*, menambah data *training*, menambah fitur kelas tidak hanya *cyberbullying* dan *non-cyberbullying* saja tapi hasil kelas *cyberbullying* bisa diekstrak menjadi kelas yang lebih spesifik lagi seperti *cyberbullying* berupa *bully*, *racism*, *sexism*, dan *attack*. Terakhir menambah algoritma untuk dibandingkan untuk melihat algoritma mana yang lebih baik.

Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM

ORIGINALITY REPORT

7%

SIMILARITY INDEX

3%

INTERNET SOURCES

5%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

- 1** Ahmad Fathan Hidayatullah, Aufa Aulia Fadila, Kiki Purnama Juwairi, Royan Abida Nayoan. **Jurnal Linguistik Komputasional (JLK), 2019**
Publication **3%**
- 2** Imam Kurniawan, Ajib Susanto. "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019", Eksplora Informatika, 2019
Publication **1%**
- 3** www.warse.org
Internet Source **1%**
- 4** id.123dok.com
Internet Source **1%**
- 5** journal2.uad.ac.id
Internet Source **<1%**
- 6** medium.com
Internet Source **<1%**
- 7** bmcbioinformatics.biomedcentral.com
Internet Source

<1%

8

managementmyblog.blogspot.com

Internet Source

<1%

9

Alejandro Garcia-Romero, Araceli Hernandez-Vitoria, Esther Millan-Cebrian, Veronica Alba-Escorihuela et al. "On the new metrics for IMRT QA verification", Medical Physics, 2016

Publication

<1%

10

journal.unpar.ac.id

Internet Source

<1%

11

tekno.kompas.com

Internet Source

<1%

12

Yaohong Ma, Hong Fan, Cheng Zhao. "Feature-Based Fusion Adversarial Recurrent Neural Networks for Text Sentiment Classification", IEEE Access, 2019

Publication

<1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On