Jurnal Khazanah Volume 14 No. 2, November 2022
E-ISSN 2745-8733
P-ISSN 1979-519X

IDENTIFIKASI KONTEN DEWASA PADA CUITAN TWITTERMENGGUNAKAN METODE BILSTM SEBAGAI UPAYA MENGATASI PENYEBARAN PORNOGRAFI UNTUK INDONESIA MAJU

Akmal Perdana Hesaputra[1], Rayhan Digo Saputra[2], Yafi Hudatama Wibowo[3]

[1,2,3] Program Studi Informatika, Fakultas Teknologi Industri,
Universitas Islam Indonesia

ABSTRAK

Penggunaan bahasa dalam kehidupan merupakan suatu hal yang sangat mendasar, termasuk pada jejaring media sosial yang sudah menjadi sesuatu yang tak terpisahkan. Namun, banyak konten di media sosial berisi informasi yang tidak bermanfaat dan diperlukan dengan beredarnya konten-konten negatif dan berbahaya, seperti konten dewasa atau pornografi. Fungsi otak akan berubah pada seseorang yang memiliki kecanduan, salah satunya kecanduankonten pornografi. Maka dari itu, konten-konten dewasa tersebut merupakan ancaman serius yang dapat membahayakan generasi muda bangsa Indonesia, terutama anak-anak dan remaja yang merupakan cikal bakal menjadi tonggak bagi kemajuan dan pembangunan bangsa. Olehkarena itu, penelitian ini bertujuan untuk membangun model klasifikasi konten dewasa menggunakan algoritma Bidirectional Long Short Term Memory (BiLSTM) Neural Network untukmengklasifikasi konten dewasa dan non-dewasa pada media sosial Twitter dengan memanfaatkan Twitter API dan library Tweepy. Berdasarkan percobaan, model terbaik diperoleh dari model BiLSTM Double layer dengan dropout yang memiliki Accuracy 98.34% danF1-Score sebesar 98.32%.

Kata kunci: Generasi Muda, Konten Dewasa, Twitter, Klasifikasi, BiLSTM

ABSTRACT

The use of language in life is a very basic thing, including in social media networks which havebecome something that cannot be separated. However, a lot of content on social media contains useless and necessary information with the circulation of negative and harmful content, such asadult content or pornography. Brain function will change in someone who has addiction, one ofwhich is pornography addiction. Therefore, adult content is a serious threat that can endanger the young generation of the Indonesian nation, especially children and adolescents who are theforerunners to the progress and development of the nation. Therefore, this study aims to build a classification model for adult content using the Bidirectional Long Short Term Memory (BiLSTM) Neural Network algorithm to classify adult and non-adult content on Twitter social media by utilizing the Twitter API and the Tweepy library. Based on the experiment, the best model is obtained from the BiLSTM Double layer model with a dropout that has an Accuracy of98.34% and an F1-Score of 98.32%.

Keywords: *Young Generation, Adult Content, Twitter, Classification, BiLSTM*

1. PENDAHULUAN

Penggunaan bahasa dalamkehidupan merupakan suatu hal yang sangat mendasar, termasuk padajejaring media sosial yang sudahmenjadi sesuatu yang tak terpisahkan. Namun, banyak konten di media sosial berisi informasi yang tidak bermanfaat dan diperlukan dengan beredarnya konten konten negatif dan berbahaya, seperti konten dewasa atau

pornografi. Konten konten negatif tersebut merupakan ancaman serius yang dapat membahayakan generasi bangsa,terutama anak-anak dan remaja. Konten konten dewasa tersebut dapat direpresentasikan dalam bentuk teks, gambar, dan video.

Pada Januari hingga September2020, Kementerian Komunikasi dan Informatika Indonesia melaporkan lebih dari 1,3 juta konten negatif telahdiposting di internet termasuk media sosial. Diantara konten negatif tersebut, pornografi merupakan yang paling banyak dengan lebih 1,06 juta konten. Keberadaan konten pornografi di media sosial merupakanyang paling banyak dan sangat mengancam yang harus diwaspadai.

McKinsey Global Institute menyatakan estimasi Indonesia akanmenjadi negara maju pada tahun2030, Gross Domestic Product (GDP) Indonesia bisa menempati urutannomor 7 dunia. Hal ini didukung dengan peningkatan kelas menengah 1 dari 45 juta orang pada tahun 2013 menjadi 135 juta orang pada tahun 2030. Namun, terdapat kendala utama yang menghalangi untukmenjadi negara maju terkait dengan kualitas Sumber Daya Manusia.Selain kualitas, karakter SDM yang unggul juga perlu dibentuk sehinggadapat berprestasi dan terhindar dari paparan konten-konten negatif seperti konten porno. Menurut survey yang Kemenkes tahun 2017, sebanyak 95.1% siswa SMP dan SMA pernah mengakses konten porno. Paparan pornografi padaanak-anak dan remaja ini dapat menyebabkan kecanduan pornografi. Kecanduan pornografi juga mengakibatkan kerusakan otak yangcukup serius yang menyerang PreFrontal Korteks (PFC), otak yangmerupakan salah satu bagian yangpaling penting karena bagian otak ini hanya dimiliki oleh manusia.

Kecanduan pornografi terhadap anak-anak dan remaja menjadi salah satu faktor penghambat kemajuan bangsa, karena hal tersebut menyebabkan perilaku menyimpang yang merusak mental dan moral generasi muda, sehingga hal ini akanberdampak kepada bibit unggul generasi muda yang ingin bersinergi memajukan pembangunan bangsa Indonesia. Beberapa penelitian terkait yaitu, penelitian oleh Izzah dkk memanfaatkan algoritma Naive Bayes Classifier dan Support Vector Machine untuk menampilkan konten pornografi di media sosial.(1) Hasil penelitian mereka menyatakan bahwa t model terbaik diperoleh dari kombinasi Support Vector Machine kata yang paling umum, dan kombinasi unigram dan bigram, dengan nilai F1-Score 91,14%.

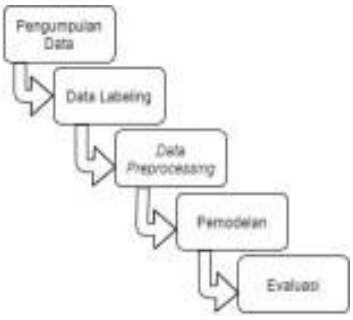
Selain itu, algoritma yang banyak digunakan oleh peneliti untuk klasifikasi teks adalah Long Short Term Memory (LSTM) Neural Network. Pitsilis dkk menggunakan LSTM untuk mengidentifikasi tweet yangmengandung rasisme, seksisme, dankonten netral.(2) Wang dkk melakukan analisis sentimen menggunakan LSTM yang dikombinasikan dengan penyisipan kata.(3) Hidayatullah dkk mempresentasikan klasifikasi konten dewasa pada data Twitter menggunakan Long Short TermMemory (LSTM) Neural Network.(4) Hasil penelitian mereka menunjukkan model terbaik diperoleh dengan menerapkan 2 lapisan LSTM dengandropout dengan akurasi 98,38% dan juga menemukan bahwa adanya dropout mempengaruhi nilai loss dan nilai akurasi. Penelitian tersebut juga mengungkapkan bahwa LSTM menunjukkan kinerja yang lebih baik dibandingkan dengan beberapa metode pembelajaran mesin tradisional, termasuk Multinomial Naïve Bayes, Logistic Regression and Support Vector Classification. Disisi lain Fadli dkk melakukan penelitian dengan membandingkan antara LSTM dan BiLSTM terhadap cyberbullying yang terjadi di twitter.(5) Penelitian tersebut membuktikan bahwa BiLSTM mendapatkan hasil accuracy sebesar 94,51% terpaut0,74% lebih baik daripada LSTM. Berdasarkan permasalahan tersebut, masifnya penyebaran konten dewasa di Indonesia sangat penting untuk diatasi. Salah satu upaya yang dapat dilakukan untuk menghindari

penyebaran konten dewasa adalah dengan memblokir teks atau kalimat yang mengandung kata-kata yang menjurus ke arah seksual. Untuk itu, kita perlu menentukan kalimat yang mengandung kata-kata yangmenjurus ke arah seksual atau tidak. Salah satu tantangannya adalah bagaimana membedakan antara kata-kata yang menjurus ke arah seksual yang digunakan dalam konteks non-dewasa. Misalnya, kata 'seks (artinya: seks) dapat digunakan baik dalam konteks seksual maupun non-seksual.

Oleh karena itu, penelitian inibertujuan untuk membangun model untuk membantu mengidentifikasi konten dewasa pada Twitter. Twitter merupakan salah satu platform media sosial yang sering digunakan untuk menyebarkan konten pornografi, termasuk teks dan gambar.(6) Selainitu, kami menggunakan algoritma Bidirectional Long Short Term Memory (BiLSTM) BiLSTM Neural Network untuk mengklasifikasikan konten dewasa dan konten non- dewasa. Sistematika penyusunan makalah ini disusun sebagai berikut: Bagian 1 menjelaskan Pendahuluan; Bagian 2 menyediakan Kajian Literatur; Bagian 3 menjelaskan Metodologi; Bagian 4 menyajikan Hasil dan Pembahasan; Bagian 5 menyajikan Kesimpulan; dan Bagian 6 sebagai Referensi dari pekerjaan kami.

2. METODE

Alur pengerjaan penelitian ini terdiri dari pengumpulan data, data labeling, data preprocessing,pemodelan, dan evaluasi. Gambar 1meninjukan alur pengerjaanpenelitian ini



Gambar 1. Alur Pengerjaan Penelitian

2.1 Pengumpulan Data

Dataset yang digunakan merupakan data dari penelitian sebelumnya yang dilakukan oleh Hidayatullah dkk yang mengklasifikasikan konten dewasa pada data Twitter menggunakan Long Short Term Memory (LSTM) Neural Network.(4) Proses pengumpulan data tersebut dilakukan dengan memanfaatkan Twitter API dan library tweepy. Pengambilan data dilakukan pada bulan April, Mei, Agustus dan September 2019 dengan data yang terkumpul sebanyak 52.338 dataset mentah. Pengambilan data berupa konten dewasa dan konten non dewasa dilakukan dengan dua cara yaitu, berdasarkan akun dan kata kunci. Untuk pengambilan data konten dewasa, dilakukan dengan cara mengidentifikasi secara manual beberapa akun Twitter yang sering memposting konten pornografi dan konten dewasa, kemudian kami memasukkan akun tersebut ke dalam daftar. Selanjutnya, kami mengumpulkan tweet berdasarkan akun Twitter yang telah dimasukkan dalam daftar. Kami juga membuat daftar kata kata porno dan seksualitas untuk digunakan sebagai kata kunci. Sedangkan untuk datasetnon-dewasa, kami mengambil data dari akun Twitter berita yang memposting informasi umum seperti akun pendidikan, pemerintahan, dan kesehatan.

2.2 Data Labelling

Pada proses data labeling kami menggunakan pseudo labelling untuk melabeli dataset. pseudo labelling adalah metode semi-diawasi yang menggunakan sejumlah kecil data berlabel untuk memberi label lebih banyak kumpulan data yang tidak berlabel. Dalam penelitian ini, labeled data yang digunakan sebanyak 1000 tweet yang terdiri dari 500 tweet konten dewasa dan 500 tweet bukan konten dewasa. Kemudian Data tersebut digunakan untukmemprediksi lbel pada 51.228 tweet yang belum dilabeli.

2.3 Data Preprocessing

Data preprocessing bertujuan untuk mengurangi dimensi data dengan menghilangkan noise dan beberapa hal

yang tidak perlu dalam dataset. Dalam penelitian ini, kami melakukan preprocessing data untuk dataset Twitter dengan mengikuti langkah-langkah preprocessing yang dilakukan oleh Hidayatullah danMa'arif [8]:

1. Menghapus URL
2. Melakukan *Casefolding*
3. Menghapus Angka
4. Menghapus karakter khusus Twitter
5. Menghilangkan tanda baca
6. Menghapus spasi ganda
7. Menghapus karakter non-ASCII
8. Menghapus emoticon
9. Normalisasi kata gaul
10. Menghapus karakter berulang

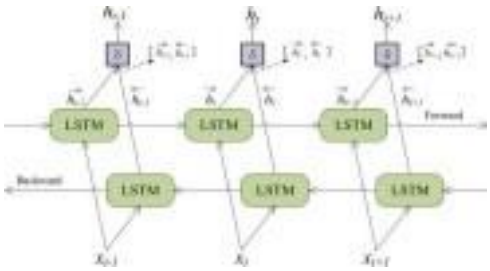
Proses ini dieksekusi menggunakan bahasa pemrograman Python pada Google Colab. Setelah itu, hasil tersebut disimpan dalam file berformat xlsx. Tabel 1 menampilkan beberapa contoh data sebelum dan sesudah dilakukan *Preprocessing*.

Tabel. 1 Perbandingan Text Sebelum dan Sesudah Dilakukan Preprocessing

Sebelum	Sesudah
RT @penggun a1: pendidikan seksual bukan berarti menyuruh kita untuk berhubungan seks kan? kenapa masih tabu aja ya https://t.co/ps 90 hkV…	pendidikan seksual bukan berarti menyuruh kita untuk berhubungan seks kan kenapa masih tabu saja iya

2.4 **Pemodelan**

Proses pemodelan dilakukan menggunakan metode Bidirectional Long Short Term Memory (BiLSTM) yang merupakan pengembangan dari algoritma Long Short-Term Memory (LSTM). Algoritma ini memiliki dua lapisan dengan proses yang saling berkebalikan arah. Lapisan bawah yang bergerak maju untuk memahami dan memproses dari kata pertama menuju kata terakhir sedangkan lapisan atas bergerak mundur memahami dan memproses dari kata terakhir menujukata pertama. Gambar 2 menunjukkan struktur dari algoritma BiLSTM.



Gambar 2. Struktur Algoritma BiLSTM

Terdapat 3 layer yang menjadi patokan dalam proses membangun model, yaitu embedding layer, BiLSTM layer, dan dense layer. Embedding layer berfungsi untuk merepresentasikan kedekatan makna tiap kata dengan mengubah data latih menjadi vektor numerik. BiLSTM layer berisi 2 LSTM layer yaitu forward LSTM dan backward LSTM sehingga gabungan tersebut akan menangkap informasi darikedua arah. Sedangkan Dense layer berfungsi sebagai output layer. BiLSTM menghitung *forward hidden sequence, backward hidden sequence*, dan kemudian digabungkan untuk menghitung *output sequence*-nya. Dapat digambarkan dengan rumus seperti yang ditunjukan pada gambar 3.

$$\vec{h}_t = H(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}})$$

$$\overleftarrow{h}_t = H(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}})$$

$$y_t = W_{\vec{h}y}\vec{h}_t + W_{\overleftarrow{h}y}\overleftarrow{h}_t + b_y$$

Gambar 3. Rumus Dasar BiLSTM

Untuk mengatasi overfitting pada model, maka diterapkan metode regularisasi yakni dropout. Dalam pembangunan model, kami menggunakan 5 *epochs* dan *32batch size* untuk membangun model klasifikasi terbaik. Penggunaan 5 *epochs* dilakukan untuk menghindari agar tidak terjadi overfitting pada model danjuga digunakan agar model dapat melakukan pembelajaran mesin. Selain itu, penggunaan 32 *batch size* dilakukan untuk memberikan*noise* atau *generalization error* yang lebih rendah. Dalam pembangunan model ini, kami menggunakan empat metodeberbeda, yaitu:

1. BiLSTM Single layer tanpadropout
2. BiLSTM Double layer tanpadropout
3. BiLSTM Single layer dengandropout
4. BiLSTM Double layer dengandropout

Dari beberapa metode tersebut akan diambil model terbaik yang akan digunakan dalam melakukan pendeteksian konten dewasa pada cuitan berbahasa Indonesia. Tabel 2 menunjukkan perbandingan hyperparameter yang digunakan dalam pembangunan model.

Tabel 2. Perbandingan Hyperparameter

Model	**LSTM Unit**	**Batch Size**	**Epoch**	**Drop out**
BiLSTM Single layer tanpa dropout	128	32	5	-
Double layer tanpa dropout	128, 128	32	5	-
BiLSTM Single layer tanpa dropout	128	32	5	0.9, 0.5
Double layer tanpa dropout	128, 128	32	5	0.9, 0.5, 0.4

2.5 Evaluasi

Evaluasi bertujuan untuk mengukur kinerja model. Dalam pekerjaan ini, kami membagi dataset kami menjadi tiga bagian yang berbeda, data pelatihan, data validasi dan data pengujian. Data pelatihan akan dilatih sebagai model pelatihan. Data validasi digunakan. untuk mengukur kinerja model selama proses pelatihan. Data pengujian digunakan untuk mengevaluasi kinerja model yang dibuat. Untuk mendapatkan kinerja terbaik, kami menggunakan rumus *Confusion Matrix* untuk mengukur kinerja dari setiap metode. Tabel 3 menunjukkan struktur dari *Confusion Matrix*. Gambar 4 menyajikan rumus *Confusion Matrix*.

Tabel 3. Confusion Matrix

Actual Value	Predicted Value	
	Negative	Positive
Negative	TN	FP
Positive	FN	TP

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2(Precision * Recall)}{Precision + Recall}$$

Gambar 4. Rumus Confusion Matrix

Keterangan:
True Positive (TP): jumlah data dari kelas positif yang benar diklasifikasikan sebagai kelas positif.
True Negative (TN): jumlah data dari kelas negatif yang benar diklasifikasikan sebagai kelas negatif.
False Positive (FP): jumlah data dari kelas positif yang salah diklasifikasikan sebagai kelas negatif.
False Negative (FN): jumlah data dari kelas negatif yang salah diklasifikasikan sebagai kelas positif.

3. HASIL DAN PEMBAHASAN

Berdasarkan pengujian model yang telah dibangun, dengan menggunakan Confusion Matrix maka diperoleh True Positive (TP), kalimat yang diprediksi (Positif) mengandung konten dewasa dan memang benar (True) mengandung konten dewasa. Penentuan True Negative (TN), kalimat yang diprediksi (Negatif) mengandung konten dewasa dan memang benar (True) tidak mengandung konten dewasa. Penentuan False Positive (FP), kalimat yang diprediksi (Positif) mengandung konten dewasa, tetapi prediksi tersebut salah (False) karena tidak mengandung konten dewasa. Selanjutnya, False Negative (FN), kalimat yang diprediksi (Negatif) mengandung konten dewasa, tetapi sebenarnya benar (True) mengandung konten dewasa. Tabel 4 menampilkan hasil dari Confusion Matrix setiap model BiLSTM.

Tabel 4. Confusion Matrix Setiap Model

BiLSTM Single layerdengan dropout

Actual Value	**Predicted Value**	
	0	1
0	8943	192
1	125	8672

BiLSTM Single layertanpa dropout

Actual Value	**Predicted Value**	
	0	1
0	8944	191
1	140	8657

BiLSTM Double layer dengan dropout		
Actual Value	**Predicted Value**	
	0	1
0	8932	203
1	95	8702

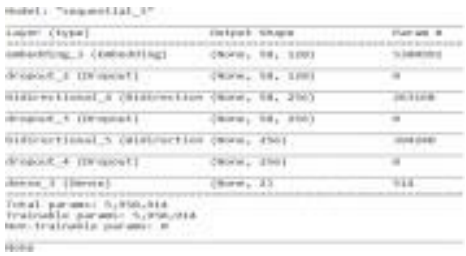
BiLSTM Double layer tanpa dropout		
Actual Value	**Predicted Value**	
	0	1
0	8959	176
1	214	8583

Dari data yang sudah didapatkan dapat dihitung Accuracy, Precision, Recall, dan juga F1-Score. Nantinya hasil dari model tersebut akan dibandingkan untuk dicari model yang terbaik. Berdasarkan hasil hitungan Accuracy, Precision, Recall, dan juga F1- Score yang didapat model terbaik adalah model yang memiliki persentase yang paling tinggi dari model lainnya. Tabel 5 menyajikan akurasi setiap model Confusion Matrix.

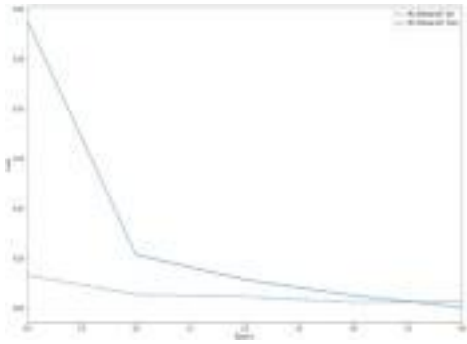
Tabel 5. Akurasi Confusion Matrix

Model	Accuracy	Precision	Recall	F1-Score
BiLSTM Single layertanpa dropout	98.15 %	97.83%	98.41%	98.12%
Double layer tanpadropout	97,83 %	97.99%	97.57%	97.78%
BiLSTM Single layertanpa dropout	98.23 %	97.83%	98.58%	98.20%
Double layer tanpadropout	98.34 %	97.72%	98.92%	**98.32%**

Dari data tersebut didapatkan bahwa BiLSTM Double layer dengan dropoutmerupakan model terbaik dalam percobaan dengan memperoleh*accuracy* 98.34% dan *F1-Score* 98.32%. Gambar 5 dan 6menampilkan summary dan grafik proses training dari model BiLSTM Double layer dengan dropout



Gambar 5. Summary BiLSTM DoubleLayer Dengan Dropout



Gambar 6. Grafik Proses Training Model BiLSTM Double Layer Dengan Dropout

Dari model tersebut dibuat *interface* sederhana untuk mengecek suatu kalimat tergolong ke dalam konten dewasa atau tidak. Gambar 7 menampilkan hasil pengujian identifikasi kalimat yangmengandung konten dewasa. Gambar 8 menampilkan hasil pengujian kalimat yang tidak mengandung konten dewasa

Gambar 7. Ui Coba Terhadap Kalimat yang merupakan Konten Dewasa

Gambar 8. Uji Coba Terhadap kalimat yang Bukan merupakan Konten Dewasa

4. KESIMPULAN

Dalam penelitian ini, kami mempresentasikan Identifikasi Konten Dewasa pada Cuitan Twitter MenggunakanMetode BiLSTM Sebagai Upaya Mengatasi Penyebaran Pornografi Untuk Indonesia Maju. Berdasarkan percobaan yang telah dilakukan, model terbaik diperoleh dari model BiLSTM Double layer dengan dropout yang memiliki Accuracy 98.34%dan F1-Score sebesar 98.32%. Modeltersebut dapat dijadikan sebagai metode dalam mengidentifikasi konten pornografiyang beredar di media sosial khususnya Twitter. Hal ini merupakan salah satu upaya yang dapat dilakukan untuk mencegah penyebaran konten dewasa sehingga kualitas Sumber Daya Manusia menjadi unggul.

DAFTAR PUSTAKA

1. N. Izzah, I. Budi and SL. Classification of pornographic content on Twitter using Support Vector Machine and Naive Bayes. 4th Int Conf Comput Technol Appl. 2018;156–60.
2. G. K. Pitsilis, H. Ramampiaro A, Langseth H. Detecting Offensive Language in Tweets Using Deep Learning. Appl Intell. 2018;48(12):4730–4742.
3. J.-H. Wang, T.-W. Liu, X. Luo A, Wang L. An LSTM Approach to Short Text Sentiment Classification with Word Embeddings. 2018 Conf Comput Linguist Speech Process ROCLING. 2018;214–23.

4. A. F. Hidayatullah, Anisa M. Hakim AAS. Adult Content Classification on Indonesian Tweets using LSTM Neural Network. Int Conf Adv Comput Sci Inf Syst. :235–240.
5. H. F. Fadli AFH. Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM. AUTOMATA. 2021;2.
6. E. Barfian, B. H. Iswanto and SMI. Twitter Pornography Multilingual Content Identification Based on Machine Learning. Procedia Comput Sci. 2017;116:129–36.