TEKNOMATIKA, Vol.12, No.02, September 2022
P-ISSN : 2087-9571, E-ISSN : 2541-335X

Model Klasifikasi Ujaran Kebencian pada Data Twitter dengan Menggunakan CNN-LSTM

HATE SPEECH CLASSIFICATION MODEL ON TWITTER DATA USING CNN-LSTM

Muhammad Ridwan[1], Ari Muzakir*[2]

[1,2]Universitas Bina Darma: Jl. Jenderal A. Yani No. 3 Palembang 30111, Indonesia
[1]Jurusan Teknik Informatika, Universitas Bina Darma, Palembang, Indonesia
[2]Program Doktor Sistem Informasi, Sekolah Pascasarjana, Universitas Diponegoro, Semarang, Indonesia
*e-mail: [1]181420079@ student.binadarma.ac.id, *[2]arimuzakir@binadarma.ac.id*

Abstrak

Ujaran kebencian sering terjadi di media sosial dikarenakan karakteristiknya yang bersifat publik dan transparan. Jika dibiarkan akan banyak timbul berbagai dampak negatif seperti diskriminasi, konflik sosial, dan bahkan genosida. Untuk menghindari hal tersebut pencegahan dengan mendeteksi ujaran kebencian harus dilakukan. Penelitian ini mengusulkan metode hybrid deep learning yang terdiri dari gabungan model Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) yang didukung oleh metode word embedding Skip-gram dan Continuous Bag of Word (CBOW) dari model Word2Vec untuk membuat model klasifikasi yang dapat bekerja pada ujaran kebencian di Twitter. Eksperimen dilakukan dengan menyetel kombinasi iterasi dan dimensi embedding pada model CNN-LSTM dengan Skip-gram dan CBOW. Hasil CNN-LSTM terbaik didapatkan dari kombinasi 30 iterasi dan 300 dimensi Skip-gram yang memperoleh nilai akurasi label setinggi 69.1% pada tahapan uji coba.

Kata kunci —ujaran kebencian, media sosial, cnn, lstm

Abstract

Hate speech often occurs on social media because of its public and transparent characteristics. If left unchecked, there will be many negative impacts such as discrimination, social conflict, and even genocide. To avoid this, prevention by detecting hate speech must be done. This study proposes a hybrid deep learning method consisting of a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models which are supported by Skip-gram and Continuous Bag of Word (CBOW) word embedding methods from the Word2Vec model to create a classification model that can work on handling hate speech on Twitter. Experiments were carried out by setting the combination of iteration and embedding dimensions on the CNN-LSTM model with Skip-gram and CBOW. The best CNN-LSTM results were obtained from a combination of 30 iterations and 300 Skip-gram dimensions which obtained a label accuracy value as high as 69.1% at the test stage.

Keywords *—hate speech, social media, cnn, lstm*

1. PENDAHULUAN

Media sosial hadir dalam berbagai bentuk dan tujuan yang dimana fungsi utamanya adalah menjadi alat penghubung bagi para penggunanya[1], akan tetapi dengan adanya kemudahan bagi setiap pengguna untuk memanfaatkan media sosial sebagai panggung

untuk berekspresi secara bebas konsekuensi yang dihasilkan dari kebebasan tersebut juga harus dipertimbangkan[2].

Dikarenakan sarana komunikasi media sosial yang bersifat publik dan transparan. Dengan karakteristik tersebut, media sosial dapat menjadi pemicu tinggi bagi masyarat untuk cenderung melakukan ujaran kebencian di kolom komentar media sosial[3]. Hal ini terutama sering terjadi di media sosial besar ternama seperti Twitter, sebuah media sosial dengan 18,45 juta pengguna[4] di tahun 2022 yang dimana hal itu menyebabkan ujaran kebencian yang terjadi di Twitter sangatlah mengkhawatirkan apabila difaktorkan jumlah penggunanya. Dinyatakan [5][6] bahwa ujaran kebencian dan bahasa kasar di media sosial harus dideteksi untuk mencegah tindakan yang dapat menimbulkan berbagai dampak negatif seperti diskriminasi, konflik sosial, dan bahkan genosida. Pencegahan dengan mendeteksi ujaran kebencian harus dilakukan agar masyarakat dan anak-anak terhindar dari mempelajari dan mengutarakan ujaran kebencian beserta bahasa-bahasa yang tidak pantas di media sosial.

Terdapat penelitian terhadap klasifikasi ujaran kebencian dan kata-kata kasar yang sudah pernah dilakukan sebelumnya[7][8][9]. Pada penelitian[7] klasifikasi ujaran kebencian dan kata-kata kasar berbahasa Indonesia diperoleh data dari media sosial Twitter. Jenis klasifikasi yang dilakukan adalah multi-label yang dimana klasifikasi berfokus terhadap sasaran, kategori, dan tingkat ujaran kebencian yang dilontarkan. Hasil terbaik diperoleh melalui machine learning dengan hasil terbaiknya menggunakan metode Random Forest Decision Tree (RFDT) dengan Label Powerset dan Term Frequency–Inverse Document Frequency (TF-IDF) untuk metode transformasi datanya. Dengan menggunakan data yang sama[8][9], kedua penelitian ini melakukan komparasi terhadap performa algoritma *machine learning* dan *deep learning* untuk mendapatkan hasil klasifikasi terbaik. Penelitian[8] menggunakan algoritma RFDT, Naive Bayes, Support Vector Machines (SVM), Bayesian Logistic Regression dan Bidirectional Long Short-Term Memory Neural Network (BiLSTM). Berhasil didapatkan hasil klasifikasi terbaik melalui kombinasi RFDT dan TF-IDF[10]. Pada penelitian[9] lebih di fokuskan komparasi antara SVM dan CNN dengan *pre-trained* DistilBERT model. Dari berbagai macam kombinasi yang dilakukan untuk menemukan hasil klasifikasi terbaik dibuktikan bahwa klasifikasi terbaik didapatkan oleh kombinasi algoritma SVM.

Pada penelitian ini, kami mengusulkan untuk melakukan klasifikasi menggunakan metode hybrid deep learning[11] yang dimana dua algoritma yaitu CNN dan LSTM digunakan untuk melakukan klasifikasi terhadap ujaran kebencian dan kata-kata kasar. Algoritma CNN-LSTM akan dikombinasikan dengan dua model Word2Vec yaitu Skip-gram dan CBOW untuk mencari hasil klasifikasi dengan tingkat akurasi terbaik terhadap ujaran kebencian dan kata-kata kasar di media sosial.

2. METODE PENELITIAN

2.1 Dataset

Dataset yang dipergunakan pada penelitian ini merupakan Twitter dataset yang berasal dari[7] dan terdiri atas 13169 tweet data pengguna. Penelitian ini akan tetap menggunakan label yang sama atas penelitian sebelumnya. Label dirumuskan dari hasil *Focus Group Discussion* (FGD) dengan staf pihak kepolisian Direktorat Tindak Pidana Siber Badan Reserse Kriminal Kepolisian Negara Republik Indonesia (Bareskrim Polri), yang bergerak dan bertanggung jawab dalam kejahatan dunia maya atau *cybercrime*. Dataset terdiri dari 12 label yang masing-masing merupakan tingkatan untuk ujaran kebencian. Contoh label dari data Twitter yang dipergunakan dalam penelitian ini terdapat pada Tabel 1.

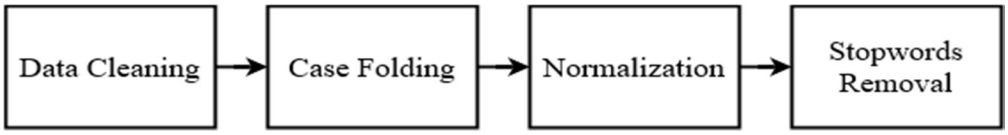
Tabel 1. Label Ujaran Kebencian

Label	Tweet	Keterangan
HS	USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah	Tweet yang menyinggung individu atau kelompok tertentu
Abusive	USER USER KO GUA KAMPRET	Tweet yang mengandung kata-kata yang bersifat kasar
HS_Individual	Pak tukang buruan pulang pak saya mau mandi, udh kaya gembel cantik ini	Tweet yang mengandung ujaran kebencian yang ditujukan terhadap seseorang atau individu
HS_Group	USER Palestina antek aseng jg tohh....	Tweet yang mengandung ujaran kebencian yang ditujukan terhadap kelompok tertentu (kelompok agama, ras, politik, dll)
HS_Religion	RT USER benci sekali dgn Umat Islam	Tweet yang mengandung ujaran kebencian yang ditujukan terhadap agama, kelompok agama, atau kepercayaan tertentu
HS_Race	Kalo seumpama orang cina di usir paksa dari indonesia gimana	Tweet yang mengandung ujaran kebencian berdasarkan penyerangan ras atau etnis
HS_Physical	USER berarti lo cacat	Tweet yang mengandung ujaran kebencian berdasarkan perbedaan fisik atau disabilitas
HS_Gender	USER Kau kan transgender Iyain aja, biar kau tenar'	Tweet yang mengandung ujaran kebencian berdasarkan gender. Biasanya berisi bahasa yang merendahkan jenis kelamin atau orientasi seksual tertentu
HS_Other	Setidaknya gw punya jari tengah buat lu, sebelom gw ukur nyali sama bacot lu	Tweet ini mengandung ujaran kebencian berupa ejekan atau hinaan yang tidak berkaitan dengan agama, ras, tipe tubuh, atau gender
HS_Weak	lebih baik jokowi mundur..	Tweet yang mengandung ujaran kebencian yang biasanya ditujukan kepada seseorang tanpa hasutan atau provokasi
HS_Moderate	Becak Warga Binjai Hendak Disita karena Gunakan Tenda Bertuliskan #2019GantiPresiden, PKS MERADANG; ; Bukti pemerintah panik dan takut?; ;	Tweet yang diyakini persilisihannya hanya terjadi di media sosial

HS_Strong	ARTINYA APA ? Pribumi sadari & lihat siapa mereka ? KECUALI KALIAN sama-sama CINA KOMUNIS / MISIONARIS abaikan tweet ini usir cina Indonesia	Tweet yang mengandung ujaran kebencian yang ditujukan kepada individu atau kelompok dengan cara yang menghasut atau provokatif

2.2 Preprocessing

Suatu tahapan penting dalam membangun model dikarenakan dengan beragamnya bentuk, jenis, dan ukuran maka kualitas dan akurasi menjadi kurang terkontrol. Postingan Twitter berisi tagar, singkatan, kesalahan ketik, dan bahasa gaul dapat menjadi pemicu untuk membahayakan kualitas dan keakuratan model[12]. Tahapan preprocessing yang dilakukan untuk memproses data Twitter ditunjukkan pada Gambar 1 dibawah ini.



Gambar 1. Preprocessing

Data dari Twitter akan diproses sesuai dengan tahapan pada Gambar 1 diatas. Hasil dari tahap *preprocessing* dipaparkan pada Tabel 2 dibawah ini.

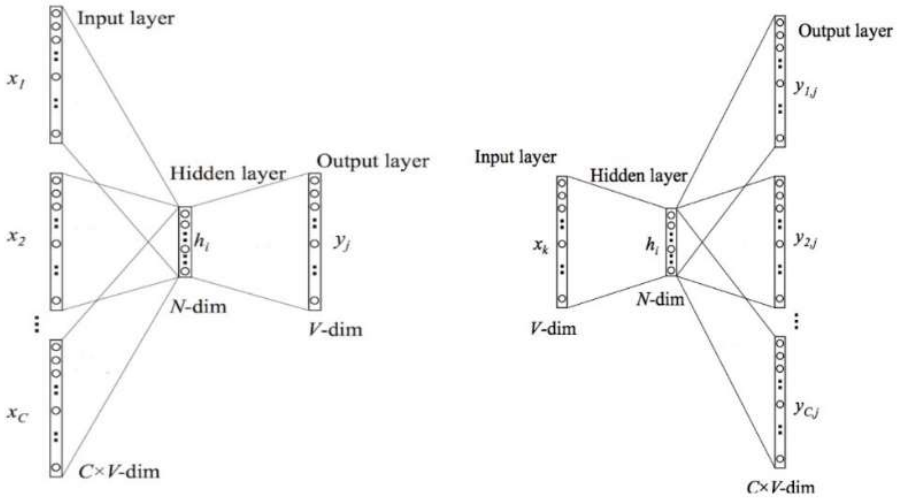
Tabel 2. Preprocessing Data Tweet

Preprocessing	Hasil Tweet
Tweet Asli	BPJS, PLN, PERTAMINA NGAKU DEFISIT SEMUA...; ; ANEHNYA PRESIDEN NGAKU HEBAT; ; KALAH SEMUA SAMA TUKANG PISANG YG OMZET TRILIUNAN
Data Cleaning	BPJS PLN PERTAMINA NGAKU DEFISIT SEMUA ANEHNYA PRESIDEN NGAKU HEBAT KALAH SEMUA SAMA TUKANG PISANG YG OMZET TRILIUNAN
Case Folding	bpjs pln pertamina ngaku defisit semua anehnya presiden ngaku hebat kalah semua sama tukang pisang yg omzet triliunan
Normalization	badan penyelenggara jaminan sosial perusahaan listrik negara pertamina mengakui defisit semua anehnya presiden mengakui hebat kalah semua sama tukang pisang yang omzet triliunan
Stopwords Removal	badan penyelenggara jaminan sosial perusahaan listrik negara pertamina mengakui defisit anehnya presiden mengakui hebat kalah tukang pisang omzet triliunan

2.3 Word Embedding

Word2vec merupakan sebuah *word embedding* model yang dapat digunakan untuk mengubah kata sehingga menjadi representasi sebuah vektor dengan panjang N, yang dimana vector tersebut tidak hanya di representasikan secara sintaksis, tetapi kata juga diwakili secara semantik. Word2Vec bekerja dengan neural network yang dimana arsitekturnya hanya terdiri dari layer *input, projection (hidden layer),* dan *output* pada rancangan arsitekturnya[12][13].

Word2Vec menyediakan dua jenis model, yaitu model Skip-gram dan CBOW. Model Skip-gram lebih dikenal sebagai cara yang efisien untuk memeriksa seberapa besar representasi vektor dalam teks yang tidak terstruktur. Arsitektur *word embedding* model Skip-gram bekerja dengan mencoba membuat prediksi pada rentang sesudah atau sebelum kata saat ini yang inputnya juga berasal dari kata saat ini, sedangkan model CBOW memprediksi kata-kata saat ini hanya berdasarkan konteks kata[15]. Gambar dari arsitektur model CBOW dan Skip-gram dapat dilihat di Gambar 2.



Gambar 2. Arsitektur CBOW(kiri) dan Skip-gram(kanan) [15]

Pada penelitian ini dilakukan pelatihan terhadap dua model di gambar atas menggunakan data Wikipedia yang bertotalkan 460 ribu artikel berbahasa indonesia dari https://dumps.wikimedia.org/idwiki/latest/idwiki-latest-pages-articles.xml.bz2.

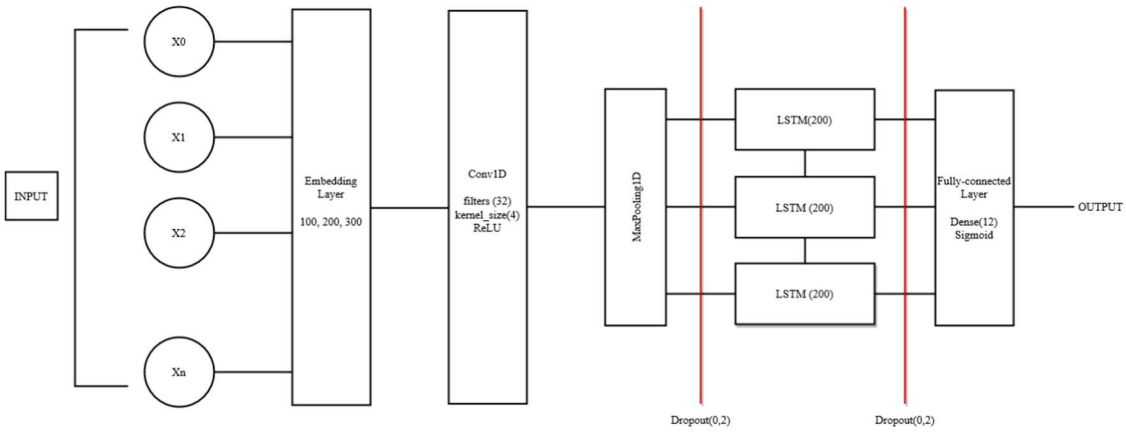
2.4 CNN-LSTM

Algoritma CNN merupakan sebuah jenis algoritma *deep learning* dikarenakan banyaknya tingkat jaringan yang dapat diimplementasikan. Arsitektur dari terbilang mirip dengan pola sel saraf otak manusia yang tersusun atas tiga jenis layer yang dinamakan *convolutional layers, pooling layers* dan *fully-connected layers*[16].

Algoritma LSTM merupakan algoritma modifikasi dari RNN (*Recurrent Neural Network*) yang diciptakan dikarenakan adanya beberapa kesulitan dalam melatih model RNN. LSTM dihadirkan untuk melengkapi kekurangan RNN yang tidak bisa memprediksi kata berdasarkan informasi lampau yang disimpan dalam jangka waktu lama sekaligus juga dapat membersihkan informasi yang tidak dianggap lagi relevan. LSTM terbilang lebih efisien dalam upaya

memproses, memprediksi, dan sekaligus mengklasifikasikan data berdasarkan urutan waktu tertentu[17].

Pada penilitian ini model CNN-LSTM bekerja atas kombinasi dari lapisan *convolutional* yang akan *input*, yang diman kemudian *outputnya* akan di utarakan ke dimensi yang lebih kecil sebelum masuk ke dalam lapisan LSTM. Tahapan akhirnya, *output* yang di keluarkan oleh lapisan LSTM akan diteruskan ke lapisan *Dense* agar hasil akhir dapat di produksi. Gambaran rangkaian proses CNN-LSTM dapat dilihat di Gambar 3 bawah ini.



Gambar 3. Arsitektur CNN-LSTM

2.5 Evaluasi

Tahapan akhir yaitu pengujian model klasifikasi yang bertujuan untuk mengetahui kemampuan dari model dengan mengukur akurasinya. Rumus untuk menghitung akurasi pada klasifikasi *multi-label* dapat dilakukan dengan persamaan berikut.

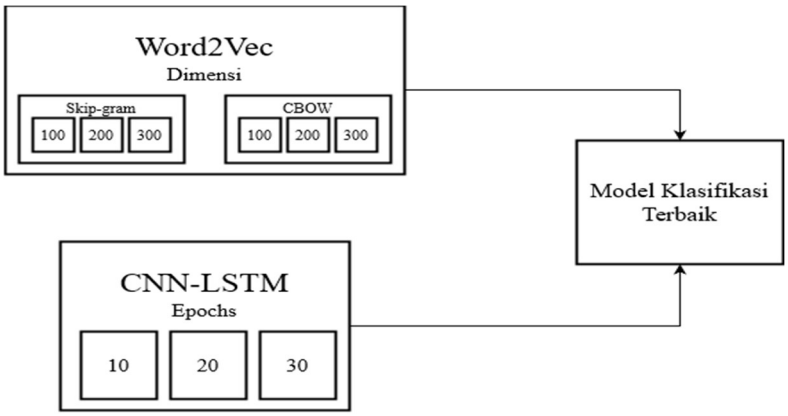
$$Accuracy = \frac{1}{D}\sum_{i=1}^{D}\left(\left|\frac{\hat{L}^{(i)} \wedge L^{(i)}}{\hat{L}^{(i)} \vee L^{(i)}}\right|\right) \times 100\% \quad (1)$$

Dalam rumus(1) nilai D merupakan jumlah dari dokumen yang terdapat di dalam dataset, $\hat{L}^{(i)}$ merupakan label prediksi untuk dokumen i, dan $L^{(i)}$ adalah label dokumen yang asli.

3. HASIL DAN PEMBAHASAN

3.1 Skenario Eksperimen

Berbagai macam skenario percobaan dilakukan pada penelitian ini yang bertujuan untuk memperoleh hasil model terbaik dalam kinerja klasifikasi ujaran kebencian di media sosial. Metode klasifikasi yang menggunakan model CNN-LSTM dan *word embedding* Skip-gram dan CBOW akan digabung dan disusun untuk dilakukan percobaan di berbagai skenario yang dimana nanti tingkat akurasi dari model akan diukur dengan menggunakan rumus(1). Gambar dari skenario eksperimen dapat dilihat di Gambar 4 bawah ini.



Gambar 4. Skenario Eksperimen

3.2 Hasil Training

Pada tahap penelitian ini dilakukan pelatihan dan validasi untuk model klasifikasi. Pelatihan dilakukan dengan menggunakan dataset Twitter yang sudah dibagi menjadi 60:20:20 untuk masing-masing *train data, validation data,* dan *test data*. Sebagaimana *train data* digunakan untuk melatih bobot dan bias *neural network*, *validation data* memiliki fungsi untuk mengevaluasi model secara langsung selama proses pelatihan berlangsung. Hasil pelatihan yang didapatkan dari percobaan sesuai dengan apa yang terdapat pada Gambar 4, menghasilkan hasil pada Tabel 3 di bawah ini.

Tabel 3. Hasil Training

Word Embedding	Dimension	Epochs	Label Acc(%)	
			Train	Val
CBOW	100	10	72.4	60.6
	100	20	73.2	61
	100	30	73.3	62.1
	200	10	74	61.2
	200	20	76	63.1
	200	30	76.1	63.3
	300	10	74.3	61.6
	300	20	77.3	63.3
	300	30	78.8	64
Skip-gram	100	10	71.2	61
	100	20	80.6	63.8
	100	30	81.8	66.1
	200	10	77.7	65.7
	200	20	82.7	66.9
	200	30	83.4	67.4
	300	10	79.6	63.2
	300	20	84.1	67.4
	300	30	85.4	68.5

Dapat diamati pada Tabel 3 diatas bahwa berbagai macam skenario yang telah dilakukan berpengaruh atas kinerja model. Performa model terlihat dengan jelas meningkat apabila digunakan jumlah *epochs* atau iterasi dan dimensi *word embedding* yang lebih besar. Nilai akurasi tertinggi dicapai sebesar 85.4% untuk *train accuracy* dan 68.5% untuk *validation accuracy* yang diperoleh dari kombinasi Skip-gram dengan 300 dimensi dan CNN-LSTM dengan 30 iterasi.

3.3 Hasil Testing

Tahapan *testing* merupakan suatu tahapan dimana model di uji dan diberi evaluasi yang tidak bias dengan menggunakan data yang tidak dikenali model selama proses pelatihan. Evaluasi dengan *test data* dilakukan agar diperoleh performa nyata dari model klasifikasi ujaran kebencian yang telah dibuat. Hasil uji coba dipaparkan pada Tabel 4 dibawah ini.

Tabel 4. Hasil Testing

Word Embedding	Dimension	Epochs	Test Acc(%)
CBOW	100	10	61.7
	100	20	62.3
	100	30	62.6
	200	10	61.9
	200	20	63.6
	200	30	63.9
	300	10	62.8
	300	20	63.5
	300	30	63.7
Skip-gram	100	10	62.2
	100	20	65.6
	100	30	66.5
	200	10	65.1
	200	20	67.3
	200	30	68
	300	10	63.7
	300	20	68.8
	300	30	69.1

Pada Tabel 4 dipaparkan hasil uji coba terhadap model-model klasifikasi yang sudah dibuat diperoleh hasil akurasi terburuk dari kombinasi CBOW dengan 100 dimensi *embedding* dan CNN-LSTM dengan 10 iterasi. Hasil terbaik diperoleh dari kombinasi Skip-gram dengan 300 dimensi dan CNN-LSTM dengan pelatihan sebanyak 30 iterasi dengan tingkat akurasi uji cobanya mencapai 69.1%. Dapat disimpulkan dengan pasti bahwa kinerja model terpengaruh atas besarnya dimensi *embedding* dan total berapa kali iterasi model pada proses pelatihan.

Pada uji coba model klasifikasi terdapat data tweet yang tidak bisa atau gagal di klasifikasi oleh model yang sudah jadi. Contoh tweet yang gagal di klasifikasi dipaparkan pada Tabel 5 di bawah ini.

Tabel 5. Permasalahan Klasifikasi

Tweet	Label Asli	Label Klasifikasi	Keterangan
salah target ahok agenda utama lengserkan jokowi	HS, HS_Individual, HS_Other, HS_Strong	HS, HS_Religion, HS_Moderate	Tweet digolongkan kedalam Ujaran Kebencian Agama dikarenakan nama "ahok" sering muncul pada tweet yang berkaitan dengan penistaan agama dan tergolong juga kedalam Ujaran Kebencian Sedang karena model gagal mendeteksi makna dari tweet karena tidak terdapat kata-kata kasar yang ekstrem seperti "mati"
sekolah agama pengukur iman	Tidak ada label	"HS_Religion"	Terdapat kata "agama" pada tweet membuat model beranggapan bahwa tweet tersebut termasuk ke dalam Ujaran Kebencian Agama

Pada Tabel 5 diatas terdapat klasifikasi yang gagal di lakukan oleh model yang sudah dibuat. Hal ini bisa terjadi dikarenakan model belum sepenuhnya dapat memahami makna dan konteks dari tweet yang gagal di klasifikasi. Terdapat beberapa faktor yang membuat model gagal dalam melakukan kinerjanya, yang pertama adanya ketidak seimbangan jenis data pada dataset dimana pada suatu label terdapat lebih banyak data, kedua adanya kemungkinan model Skip-gram dan CBOW tidak atau salah mengenali konteks dari kata dimana hal ini bisa terjadi karena pada penelitian ini model Skip-gram dan CBOW dibangun menggunakan data Wikipedia sehingga tidak dapat memahami konteks data Twitter dikarenakan banyaknya kata-kata non-standar.

4. KESIMPULAN

Pada penelitian ini diperoleh kesimpulan mengenai model klasifikasi terbaik untuk ujaran kebencian di media sosial Twitter. Model yang dibangun dengan kombinasi CNN-LSTM dan Skip-gram sebagai *word embedding*-nya berhasil memperoleh nilai akurasi pada tes uji coba terbaik dengan tingkat akurasi sebesar 69.1%. Pada tes uji coba juga didapatkan beberapa permasalahan dalam klasifikasi, yang dimana untuk sebagian tweet yang ada terdapat *test data* tidak dapat diklasifikasikan dengan baik. Hal ini disebabkan oleh ketidakseimbangan data dan model Word2Vec yang dibangun hanya dengan menggunakan data artikel yang diambil dari Wikipedia, sehingga model mengalami kesulitan untuk memahami konteks dalam tweet dikarenakan banyaknya terdapat bahasa non-standar yang digunakan.

Rekomendasi yang dapat diberikan untuk studi kedepannya adalah untuk memperkaya dataset dikarenakan beberapa label dalam dataset ini dapat terbilang tidak seimbang dengan satu sama lain. Pelatihan terhadap model Word2Vec juga disarankan untuk ditambahkan dengan data Twitter untuk memperluas semantik model agar model dapat lebih mudah memahami konteks dari data Twitter. Untuk kedepannya juga dapat diterapkan metode *deep neural networks* dengan penyetelan *hyperparameter* untuk membangun model yang lebih kompleks.

DAFTAR PUSTAKA

[1] J. C. Pereira-Kohatsu, L. Quijano-Sánchez, F. Liberatore, dan M. Camacho-Collados, “Detecting and monitoring hate speech in Twitter,” *Sensors*, vol. 19, no. 21, hal. 4654, 2019.

[2] A. C. Sari, R. Hartina, R. Awalia, H. Irianti, dan N. Ainun, "Komunikasi dan media sosial," *J. Messenger*, vol. 3, no. 2, hal. 69, 2018.
[3] D. J. Ningrum, S. Suryadi, dan D. E. C. Wardhana, "Kajian ujaran kebencian di media sosial," *J. Ilm. Korpus*, vol. 2, no. 3, hal. 241–252, 2018.
[4] M. A. Rizaty, "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022," *dataindonesia.id*, 2022.
[5] Triyanto, A. Meyria, S. Aisah, dan B. Abdilah, *Buku Saku HAM Satuan Sabhara*. 2016.
[6] Y. Chen, Y. Zhou, S. Zhu, dan H. Xu, "Detecting offensive language in social media to protect adolescent online safety," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, 2012, hal. 71–80.
[7] M. O. Ibrohim dan I. Budi, "Multi-label hate speech and abusive language detection in Indonesian twitter," in *Proceedings of the Third Workshop on Abusive Language Online*, 2019, hal. 46–57.
[8] R. Hendrawan dan S. Al Faraby, "Multilabel classification of hate speech and abusive words on Indonesian Twitter social media," in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, 2020, hal. 1–7.
[9] K. M. Hana, S. Al Faraby, dan A. Bramantoro, "Multi-label classification of indonesian hate speech on twitter using support vector machines," in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, 2020, hal. 1–7.
[10] A. Muzakir, K. Adi, dan R. Kusumaningrum, "Classification of Hate Speech Language Detection on Social Media: Preliminary Study for Improvement," in *International Conference on Networking, Intelligent Systems and Security*, 2023, hal. 146–156.
[11] A. Srivastava, V. Singh, dan G. S. Drall, "Sentiment analysis of twitter data: A hybrid approach," *Int. J. Healthc. Inf. Syst. Informatics*, vol. 14, no. 2, hal. 1–16, 2019.
[12] B. Billal, A. Fonseca, dan F. Sadat, "Efficient natural language pre-processing for analyzing large data sets," in *2016 IEEE International Conference on Big Data (Big Data)*, 2016, hal. 3864–3871.
[13] Y. Goldberg dan O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method," *arXiv Prepr. arXiv1402.3722*, 2014.
[14] T. Mikolov, K. Chen, G. Corrado, dan J. Dean, "Efficient estimation of word representations in vector space," *arXiv Prepr. arXiv1301.3781*, 2013.
[15] R. P. Nawangsari, R. Kusumaningrum, dan A. Wibowo, "Word2vec for Indonesian sentiment analysis towards hotel reviews: An evaluation study," *Procedia Comput. Sci.*, vol. 157, hal. 360–366, 2019.
[16] K. O'Shea dan R. Nash, "An introduction to convolutional neural networks," *arXiv Prepr. arXiv1511.08458*, 2015.
[17] M. Sundermeyer, R. Schlüter, dan H. Ney, "LSTM neural networks for language modeling," 2012.