



## Deteksi Tingkat Keparahan Ujaran Kebencian Menggunakan Bi-LSTM pada Teks Bahasa Indonesia

Fahrul Setiawan<sup>1)</sup>, Agung Wahyudi<sup>2)</sup>, Nur Ahlina Febriyanti<sup>3)</sup>

Program Studi Teknik Informatika – Universitas 45 Surabaya

Email: fahrulsetiawan19@gmail.com, [agungwyudi@gmail.com](mailto:agungwyudi@gmail.com), [nurahlinaf@gmail.com](mailto:nurahlinaf@gmail.com)

### Abstrak

Ujaran kebencian di media sosial merupakan fenomena yang semakin mengkhawatirkan, terutama di Indonesia yang memiliki keragaman suku, agama, ras, dan antargolongan. Identifikasi tingkat keparahan ujaran kebencian (weak, moderate, strong) sangat penting untuk menangani potensi konflik secara efektif dan proporsional. Penanganan yang tepat pada setiap level dapat mencegah eskalasi konflik dan tindak kekerasan, serta membantu dalam moderasi konten yang lebih akurat di platform media sosial. Pendekatan yang lebih rinci ini memberikan dasar yang lebih kuat bagi pembuat kebijakan dan penegak hukum dalam menanggapi ancaman berdasarkan tingkat keparahannya.

Penelitian ini menggunakan dataset berbahasa Indonesia yang terdiri dari 13.169 entri dengan anotasi untuk 12 label, yang kemudian difokuskan pada tiga label tingkat keparahan ujaran kebencian dan satu label baru untuk konten bersih. Data tidak seimbang diatasi melalui oversampling, menghasilkan 30.432 entri yang seimbang. Proses pra-pemrosesan mencakup case folding, pembersihan data, stemming, dan tokenisasi. Model BiLSTM dibangun menggunakan TensorFlow dengan tujuh lapisan, termasuk embedding, BiLSTM, dan dense layers. Hyperparameter tuning dilakukan untuk menentukan konfigurasi terbaik, yang ditemukan pada 30 epochs, 20 unit BiLSTM, dan batch size 128. Model mencapai akurasi keseluruhan 93,14%, dengan performa terbaik pada kategori strong hate speech (precision 99,35%, recall 100%, F1-score 99,67%).

Model yang dihasilkan memiliki potensi aplikasi praktis dalam moderasi konten di platform media sosial dan mendukung penegakan hukum dalam memantau ujaran kebencian. Disarankan untuk penelitian selanjutnya, eksplorasi dilakukan terhadap potensi penerapan model ini di platform web atau seluler untuk analisis real-time. Ini akan meningkatkan kapasitas deteksi dan mitigasi ujaran kebencian secara langsung.

**Kata Kunci:** ujaran kebencian, twitter, Bi-LSTM, multi label, moderasi konten

### PENDAHULUAN

Hate Speech (HS) secara harfiah berarti "ungkapan kebencian" dan didefinisikan sebagai ungkapan yang menyerang seseorang atau kelompok berdasarkan ras, agama, gender, atau orientasi seksual. Dalam konteks sosiologi masyarakat Indonesia, HS mencakup ungkapan dan penyiaran kebencian yang ditujukan kepada individu, kelompok, atau lembaga berdasarkan berbagai faktor seperti agama, kepercayaan, aliran, etnis, ras, golongan, gender, orientasi seksual, dan aspek lainnya yang dapat memicu kemarahan publik. Istilah "Ujaran Kebencian" digunakan dalam Surat Edaran Kapolri Nomor: SE/6/X/2015 sebagai terjemahan resmi dari "Hate Speech". [1].

Saat ini, kebebasan berbicara telah berkembang pesat hingga seringkali mengarah pada ujaran kebencian, terutama di dunia maya seperti media sosial, di mana setiap orang dapat menyampaikan pendapatnya secara bebas tanpa adanya batasan. [2] Berdasarkan hasil FGD, ujaran kebencian dibagi menjadi tiga tingkatan: lemah, sedang, dan kuat. Ujaran kebencian lemah adalah makian atau fitnah yang ditujukan pada individu tanpa memprovokasi konflik terbuka, dan tidak diprioritaskan oleh pihak berwenang jika tidak dilaporkan oleh korban.

Ujaran kebencian sedang mencakup makian, fitnah, stereotip, atau label yang ditujukan pada kelompok tanpa memprovokasi konflik terbuka, dan biasanya terbatas pada konflik di media sosial. Ujaran kebencian kuat mencakup makian, fitnah, stereotip, atau label yang ditujukan pada individu atau kelompok dengan provokasi untuk konflik terbuka, dan perlu segera diselesaikan karena dapat mengundang konflik luas dan kerusakan fisik di dunia nyata. [3] Mengidentifikasi level ujaran kebencian sangat penting karena membantu dalam menentukan respons yang sesuai dari pihak berwenang dan masyarakat. Pemisahan ini memungkinkan penegakan hukum yang lebih efisien, di mana pihak berwenang dapat fokus pada kasus-kasus yang memerlukan tindakan segera, seperti ujaran kebencian kuat yang berpotensi memicu konflik terbuka dan kerusakan fisik. Lebih jauh lagi, dari perspektif teknis, pengklasifikasian level ujaran kebencian dapat meningkatkan akurasi sistem deteksi otomatis. [4] Dengan memfokuskan deteksi pada level keparahan, teknologi ini dapat memberikan hasil yang lebih akurat dan dapat diandalkan, membantu dalam pemantauan dan mitigasi ujaran kebencian di platform media sosial. Dalam konteks pengembangan sistem atau aplikasi, pemisahan ini memungkinkan desain fitur yang lebih canggih untuk memantau, mendeteksi, dan menanggapi ujaran kebencian secara real-time. Misalnya, aplikasi dapat dikonfigurasi untuk mengeluarkan peringatan atau mengambil tindakan moderasi otomatis berdasarkan tingkat keparahan ujaran kebencian yang terdeteksi, sehingga dapat mencegah eskalasi konflik dan menjaga keamanan serta kenyamanan pengguna. Penelitian sebelumnya mengenai deteksi ujaran kebencian di Twitter Indonesia, penggunaan kombinasi fitur word unigram, Random Forest Decision Tree (RFDT), dan Label Powerset (LP) telah menunjukkan hasil yang cukup baik dengan akurasi rata-rata sebesar 77,36% untuk klasifikasi multi-label tanpa mengidentifikasi target, kategori, dan tingkat ujaran kebencian. Namun, akurasi menurun menjadi 66,12% saat mencakup identifikasi target, kategori, dan tingkat ujaran kebencian. [3]. Pada penelitian lain menggunakan metode Bi-LSTM untuk klasifikasi ujaran kebencian, akurasi terbaik yang diperoleh mencapai 97,66%. Namun, dalam penelitian tersebut, semua tingkat keparahan ujaran kebencian digabungkan menjadi satu kelas. Penggabungan ini menghasilkan satu label umum untuk semua bentuk ujaran kebencian tanpa memisahkan level keparahan, sehingga model hanya perlu membedakan antara ujaran kebencian dan bukan ujaran kebencian.. [5] Penelitian ini akan berfokus untuk deteksi pada level ujaran kebencian, yang terdiri

dari tiga tingkatan: weak, moderate, dan strong, dengan mengabaikan label lainnya seperti target dan kategori. Dengan memfokuskan hanya pada klasifikasi level ujaran kebencian (weak, moderate, dan strong), diharapkan dapat meningkatkan akurasi deteksi dibandingkan dengan pendekatan sebelumnya yang mencakup banyak label. Penelitian ini menggunakan metode Bi-LSTM (Bidirectional Long Short-Term Memory) yang mampu menangkap konteks dari kedua arah dalam urutan teks. Metode ini diharapkan memberikan pemahaman yang lebih mendalam tentang konteks ujaran kebencian dan meningkatkan performa klasifikasi pada setiap level keparahan.

## METODE

### Sumber Data

Data yang digunakan berasal penelitian sebelumnya [3] yang diunduh dari Kaggle yang berisi data teks Bahasa Indonesia dengan total 13.169 entri disertai dengan anotasi 0 dan 1 pada setiap label yang menandakan keberadaan ujaran kebencian atau tidak. Pada data juga memiliki 12 label yang terdiri dari hate speech, abusive, individual, group, race, religion, physical, gender, others, strong, moderate, and weak. Penelitian kali ini akan fokus kepada level dari ujaran kebencian saja, dari 12 label akan digunakan hanya 3 label saja yaitu strong, moderate, weak yang mewakili level ujaran kebencian. Juga akan ditambah 1 label baru yang mewakili ketiadaan ujaran kebencian yang akan bernilai 1 jika ketiga label lain bernilai 0.

Tabel 1. Sampel Dataset

Tweet Text	HS Weak	HS Moderate	HS Strong	Clean
Adik tengok sekeliling. Macam mana adik nak ca...	0	0	0	1
USER USER Makannya w suruh baca APBN ..biar ga...	1	0	0	0
Presiden Joko Widodo (Jokowi) masuk daftar The...	0	0	0	1
USER Pak Recep..... anda salah, itu gu...	1	0	0	0

### Pra-pemrosesan Data

Tahapan ini meliputi transformasi data mentah menjadi bentuk yang lebih mudah dipahami. Prosesnya melibatkan perubahan huruf menjadi huruf kecil, pembersihan data dari link, simbol, nomor, username, normalisasi kata-kata, stemming, dan tokenisasi teks. [6] Langkah-langkah yang dapat dilakukan pada text preprocessing meliputi:

#### 1) Case Folding

Semua data dalam dataset masih belum memiliki keseragaman huruf yang digunakan. Sebagian menggunakan huruf besar dan lainnya menggunakan huruf kecil. Untuk itu diperlukan keseragaman bentuk huruf yang digunakan supaya data yang digunakan konsisten. Proses Case Folding ini mengubah semua bentuk huruf pada data menjadi huruf kecil (lowercase).

Tabel 2. Case folding

Original	Setelah Case Folding
USER USER AKU I TU AKU\n\nKU TA U MATAMU SIPIT TAPI DILIAT DARI MANA ITU AKU'	user user aku itu aku\n\nku tau matamu sipit tapi diliat dari mana itu aku'

#### 2) Pembersihan Teks

Pembersihan data digunakan untuk membersihkan data yang tidak diperlukan. Pembersihan akan dilakukan pada URL, username, simbol, nomor, hastag dll, dalam data teks. Jika data sudah bersih dari data yang tidak perlu akan memudahkan saat pelatihan dan meningkatkan akurasi model.

Tabel 3. Pembersihan Teks

Original	Setelah Pembersihan
user user aku itu aku\ n\nku tau matamu sip it tapi diliat dari man a itu aku'	aku itu aku ku tau mat amu sipit tapi diliat da ri mana itu aku

#### 3) Stemming

Stemming yaitu mengubah kata-kata dalam Bahasa Indonesia menjadi bentuk asal mereka. Pada proses

stemming akan menggunakan data dari Pustaka python yaitu Sastrawi.

Tabel 4. Stemming

Original	Setelah Stemming
provokasi mayat politi sasi agama penyebab kekalahan pilkada dki beginilah cara cebong mendeskripsikan keka lahan junjungannya fy i ahog blm pernah ikut pemilihan apapun kec jd wakil dan dia bukan lah etnis mayoritas	provokasi mayat politisasi agama sebab kalah pilih kepala daerah daerah khusus ibukota begini cara cebong deskripsi kalah jungung for your information ahok belum pernah ikut pilih apa camat jadi wakil dan dia bukan etnis mayoritas

#### 4) Tokenisasi

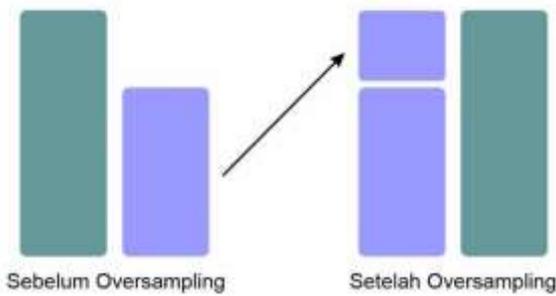
Tokenisasi memisahkan kata dan frasa menjadi bentuk token, proses ini untuk memudahkan untuk Analisa data dalam teks.

Tabel 5. Tokenisasi

Original	Setelah tokenisasi
kemarin di hujat sekarang harap dasar para bani kampret	[446, 5, 1625, 91, 555, 115, 125, 170, 131]

#### Penyeimbangan Data

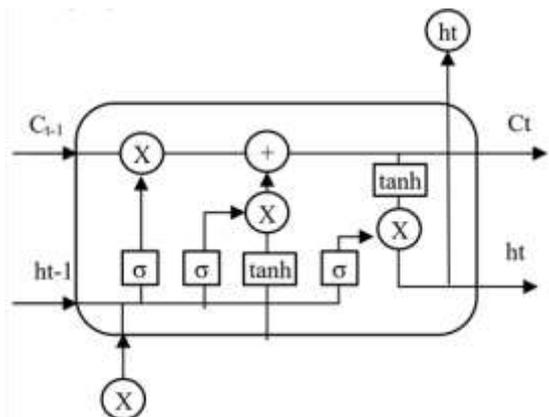
Dalam banyak kasus nyata, sering kali distribusi data antara kelas-kelas dalam sebuah dataset tidak seragam, yang berarti setidaknya satu kelas data memiliki jumlah sampel yang lebih sedikit (kelas minoritas) daripada kelas lainnya (kelas mayoritas) [7]. Salah satu cara yang efektif untuk menangani masalah ini adalah dengan oversampling. Oversampling secara acak menambah jumlah sampel dari kelas minoritas untuk menyamakan jumlahnya dengan kelas mayoritas atau untuk meningkatkan jumlah sampel secara keseluruhan. [8]



Gambar 1. Ilustrasi Oversampling

### Algoritma LSTM

Algoritma Long Short-Term Memory (LSTM) merupakan salah satu jenis arsitektur dalam Recurrent Neural Network (RNN) yang memiliki kemampuan untuk mengingat informasi jangka panjang dan mengatasi keterbatasan vanishing gradient. Arsitektur model Long Short-Term Memory (LSTM) terdiri dari beberapa komponen utama yaitu Input Gate, Forget Gate, Cell State, dan Output Gate. Input Gate menentukan informasi baru mana yang akan disimpan dalam sel memori. Forget Gate menentukan informasi mana dari sel memori yang harus dibuang atau dipertahankan. Cell State mewakili memori jangka panjang dari jaringan. Output Gate menentukan bagian dari sel memori yang akan dioutput pada langkah selanjutnya. [9]

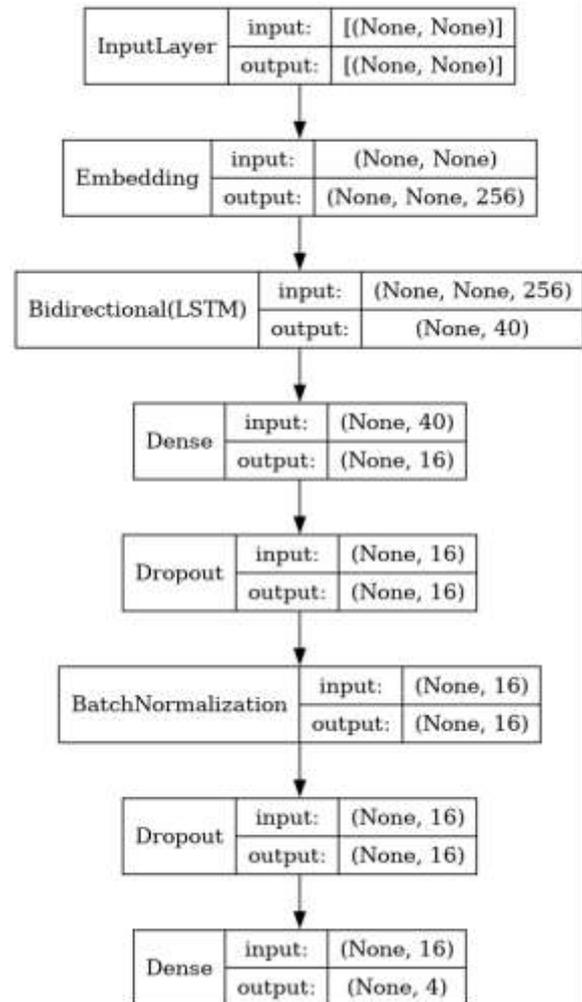


Gambar 2. Arsitektur LSTM

### Pelatihan Model

Model BiLSTM akan dibangun menggunakan framework deep learning TensorFlow. Struktur model terdiri dari beberapa lapisan, termasuk lapisan

embedding, LSTM, dan jaringan saraf yang terhubung. Lapisan embedding berfungsi sebagai penghubung yang menerima vektor hasil tokenisasi. Di ujung arsitektur, terdapat lapisan jaringan saraf terhubung yang bertindak sebagai output dari model. Lapisan ini sesuai dengan jumlah label yang digunakan dalam penelitian. Hasilnya, model akan menghasilkan output berupa vector dengan beberapa elemen, di mana setiap elemen merepresentasikan nilai dari masing-masing label.



Gambar 3. Arsitektur Model

### Evaluasi Model

Kinerja suatu algoritma sering kali dievaluasi menggunakan confusion matrix yang dilengkapi dengan classification report. Confusion matrix terdiri dari baris dan kolom di mana setiap baris mewakili kelas sebenarnya, dan setiap kolom mewakili kelas yang diprediksi [10]. Terdapat empat kombinasi yang penting, yaitu:

penelitian, objek penelitian, bahan dan alat utama, tempat, sumber data, teknik pengumpulan data, dan analisis data yang secara nyata dilakukan peneliti.

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

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

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

$$F1 - Score = \frac{2 \times (precision \times recall)}{recall + precision}$$

Dimana:

- True Positive (TP): Jumlah observasi yang secara tepat diprediksi sebagai positif oleh model.
- False Positive (FP): Jumlah observasi yang seharusnya negatif tetapi diprediksi sebagai positif oleh model.
- True Negative (TN): Jumlah observasi yang secara tepat diprediksi sebagai negatif oleh model.
- False Negative (FN): Jumlah observasi yang seharusnya positif tetapi diprediksi sebagai negatif oleh model.

## HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset yang memiliki 13.169 data, jumlah data pada masing-masing label weak, moderate, strong, clean yaitu 3.383, 1.705, 473, 7.608, ini menunjukkan data tersebut tidak seimbang. Untuk mengatasi masalah tersebut dilakukan Oversampling untuk menyeimbangkan data sehingga jumlah data setelah di Oversampling berjumlah 30.432 dengan jumlah data pada masing-masing label sama yaitu 7.608.

Sebelum diproses lebih lanjut data terlebih dahulu dilakukan pra-pemrosesan. Setelah data bersih selanjutnya dataset dibagi menjadi 2 bagian, bagian pertama data pelatihan berjumlah 80% (24.345 data), lalu data kedua data testing berjumlah 20% (6.087 data). Selanjutnya dilakukan Tokenisasi pada data agar dapat diproses oleh model.

Penelitian ini menggunakan arsitektur model berjumlah 7 layer, layer pertama yaitu layer embedding dengan nilai input 13.000 dan output 256, layer kedua yaitu BiLSTM dengan 10 unit dan aktivasi 'tanh', layer ketiga yaitu 'dense' dengan 16 unit, layer

keempat yaitu 'dropout' dengan rate 0.3, layer kelima yaitu batch normalization, keenam yaitu 'dropout' dengan rate 0.5, layer terakhir yaitu 'dense' dengan 4 unit dan aktivasi 'softmax'.

Proses modeling dilakukan tuning hyperparameter untuk menghasilkan arsitektur model terbaik. Pengaturan awal akan menggunakan epochs = 10, BiLSTM unit = 10, batch size = 32.

Hyperparameter pertama yang dituning yaitu epochs, eksperimen ini bertujuan untuk melihat berapa banyak iterasi yang diperlukan model untuk mencapai konvergen.

Tabel 6. Tuning Epoch

Epochs	Accuracy (%)
10	82.41
20	88.31
30	89.14
40	84.49

Eksperimen menunjukkan bahwa dengan meningkatnya jumlah epochs, akurasi model juga meningkat hingga mencapai titik optimal pada 30 epochs dengan rata-rata akurasi 89,14%. Setelah 30 epochs, akurasi menurun, menunjukkan bahwa model mulai mengalami overfitting pada data pelatihan. Sehingga diputuskan bahwa performa model terbaik ketika menggunakan epochs 30.

Hiperparameter kedua yang dituning yaitu jumlah unit dalam layer BiLSTM. Eksperimen ini untuk mencari jumlah unit paling optimal untuk diterapkan pada layer BiLSTM. Jumlah yang akan dicoba yaitu 10,10,30,40, dan 50.

Tabel 7. Tuning LSTM Unit

BiLSTM Units	Accuracy (%)
10	85.96
20	88.79
30	81.57
40	87.53
50	82.24

Eksperimen dengan berbagai jumlah unit BiLSTM menunjukkan bahwa 20 unit memberikan hasil terbaik

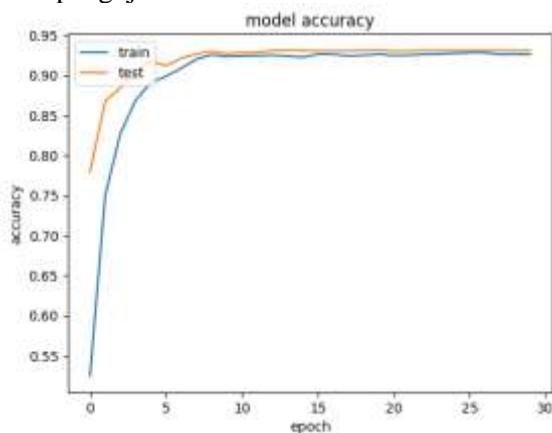
dengan rata-rata akurasi 88,79%. Penambahan jumlah unit tidak selalu berbanding lurus dengan peningkatan akurasi, dan dalam beberapa kasus, dapat menyebabkan overfitting atau peningkatan kompleksitas model tanpa manfaat tambahan yang signifikan. Sehingga unit yang akan digunakan yaitu 20 unit.

Hiperparameter ketiga yang dituning yaitu batch size, eksperimen ini menggunakan 8 variasi: 32,64,96,128,160,192,224,256. Pengujian dengan berbagai ukuran batch menunjukkan bahwa batch size 128 memberikan hasil terbaik dengan rata-rata akurasi 90,18%. Ukuran batch yang lebih besar atau lebih kecil menunjukkan penurunan performa, yang mungkin disebabkan oleh keseimbangan antara kecepatan konvergensi dan stabilitas gradien selama pelatihan., sehingga batch 128 akan digunakan pada model.

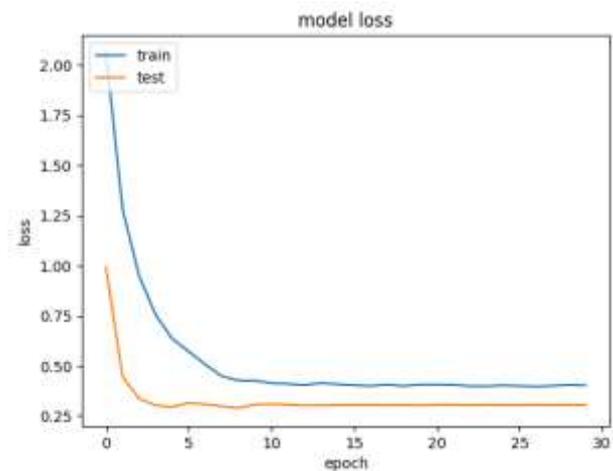
Tabel 8. Tuning Batch size

Batch Size	Accuracy (%)
64	89.54
96	89.91
128	90.18
160	88.12
192	88.11
224	89.09
256	84.59

Setelah proses tuning hyperparameter, model akhir dengan konfigurasi terbaik dievaluasi menggunakan data pengujian. Berikut hasil evaluasi akhir:



Gambar 4. Akurasi model



Gambar 5. Model loss

Tabel 9. Hasil Evaluasi

Tingkat Ujaran Kebencian	Precision	Recall	F1-Score	Akurasi
Weak	85,91%	92,58%	89,12%	93,14%
Moderate	92,87%	98,02%	95,37%	93,14%
Strong	99,35%	100%	99,67%	93,14%
Clean	94,78%	81,63%	87,72%	93,14%

Hasil evaluasi menunjukkan bahwa model BiLSTM mampu memberikan performa yang sangat baik dalam klasifikasi ujaran kebencian pada berbagai tingkat keparahan. Untuk kategori weak hate speech, model menunjukkan precision sebesar 85,91% dan recall 92,58%, menghasilkan F1-score sebesar 89,12% dengan akurasi keseluruhan 93,14%. Pada kategori moderate hate speech, precision mencapai 92,87% dan recall 98,02%, dengan F1-score 95,37% dan akurasi 93,14%. Kategori strong hate speech menunjukkan performa terbaik dengan precision 99,35%, recall 100%, dan F1-score 99,67%, serta akurasi 93,14%. Untuk kategori clean (non-hate speech), model memiliki precision 94,78%, recall 81,63%, dan F1-score 87,72%, dengan akurasi keseluruhan yang sama yaitu 93,14%. Meskipun demikian, recall yang lebih rendah pada kategori clean menunjukkan bahwa model masih memiliki tantangan dalam mengidentifikasi semua teks yang benar-benar bersih dari ujaran kebencian.

Keterbatasan dan tantangan dalam penelitian ini meliputi representasi kata di mana masalah kata-kata yang tidak ada dalam data pelatihan masih menjadi tantangan, sehingga penambahan word embedding dapat membantu meningkatkan representasi kata dan performa model. Nuansa bahasa juga menjadi tantangan karena konteks dan nuansa dalam bahasa alami bisa sangat kompleks, dan meskipun BiLSTM memberikan hasil yang baik, pendekatan seperti transformer-based models (misalnya BERT) mungkin diperlukan untuk menangkap konteks yang lebih luas. Dalam aplikasi praktis, model ini dapat digunakan untuk moderasi konten di platform media sosial, membantu mendeteksi dan mengklasifikasikan ujaran kebencian secara otomatis, serta dapat digunakan oleh lembaga penegak hukum untuk memantau dan menganalisis ujaran kebencian, membantu pencegahan konflik dan tindak kekerasan. Penelitian ini menunjukkan bahwa penggunaan BiLSTM dengan tuning hyperparameter yang tepat dapat meningkatkan akurasi deteksi ujaran kebencian secara signifikan, memberikan dasar yang kuat untuk pengembangan lebih lanjut dan aplikasi praktis dalam deteksi ujaran kebencian.

## **PENUTUP**

### **Kesimpulan**

Penelitian ini berhasil menunjukkan bahwa penggunaan model Bidirectional Long Short-Term Memory (BiLSTM) dapat memberikan performa yang sangat baik dalam klasifikasi ujaran kebencian pada berbagai tingkat keparahan. Dataset yang digunakan dalam penelitian ini terdiri dari 13.169 data, yang kemudian diatasi ketidakseimbangannya melalui metode oversampling sehingga data menjadi seimbang dengan total 30.432 entri. Setelah melalui tahapan pra-pemrosesan dan tokenisasi, data dibagi menjadi data pelatihan dan pengujian. Dengan arsitektur model yang terdiri dari 7 layer dan melalui berbagai eksperimen tuning hyperparameter, model mencapai performa optimal pada epochs 30, unit BiLSTM sebanyak 20, dan batch size 128. Hasil evaluasi menunjukkan bahwa model BiLSTM mampu mendeteksi ujaran kebencian dengan akurasi keseluruhan sebesar 93,14%. Pada kategori weak hate speech, model mencapai precision 85,91%, recall 92,58%, dan F1-score 89,12%. Untuk kategori moderate hate speech, model memperoleh precision 92,87%, recall 98,02%, dan F1-score 95,37%. Kategori strong hate speech menunjukkan performa

terbaik dengan precision 99,35%, recall 100%, dan F1-score 99,67%. Sedangkan pada kategori clean (non-hate speech), model mencapai precision 94,78%, recall 81,63%, dan F1-score 87,72%.

Penelitian ini masih menghadapi tantangan dalam representasi kata-kata yang tidak ada dalam data pelatihan dan kompleksitas konteks serta nuansa dalam bahasa alami. Oleh karena itu, penambahan embedding dan penggunaan pendekatan model berbasis transformer seperti BERT dapat dipertimbangkan untuk penelitian selanjutnya. Model ini memiliki potensi aplikasi praktis dalam moderasi konten di platform media sosial dan penegakan hukum untuk memantau dan menganalisis ujaran kebencian. Penelitian ini memberikan dasar yang kuat untuk pengembangan lebih lanjut dan aplikasi praktis dalam deteksi ujaran kebencian, dengan harapan dapat membantu mencegah konflik dan tindak kekerasan di masyarakat.

## **DAFTAR PUSTAKA**

1. N. Umar. 2019. *Jihad Melawan Religious Hate Speech*. PT Elex Media Komputindo.
2. M. A. F. R. S. P. Muhammad Mishbahul Munir. 2018. "Implementasi Metode Backpropagation Neural Network berbasis Lexicon Based Features dan Bag of Words Untuk Identifikasi Ujaran Kebencian Pada Twitter" hlm. 3182-3191.
3. I. B. M. O. Ibrohim. 2019. "Multi-label Hate Speech and Abusive Language," hlm. 46-57.
4. A. N. P. Rotua Elfrida. 2023. "Hate Speech On Social Media: A Case Study Of Blasphemy In Indonesian Context" hlm. 433-440.
5. D. H. F. S. H. Aditya Perwira Joan Dwitama. 2023. "Indonesian Hate Speech Detection Using Bidirectional Long Short-Term Memory (Bi-LSTM)" hlm. 302-309.
6. D. Soyusiawaty. 2023. *Pemrosesan Bahasa Alami*.
7. Y. A. N. S. N. R. M. A. Jafar Tanha. 2020. "Boosting methods for multi-class imbalanced data classification: an experimental review", *Journal of Big Data*, hlm. 47.
8. S. B. B. A. U. R. H. B. Ashhadul Islam. 2022. "KNNOR: An oversampling technique for imbalanced datasets", *Applied Soft Computing*, hlm. 18.



9. I. C. Y. Timothy Bastian Sianturi. 2023. "Penerapan Algoritma Long Short-Term Memory (LSTM) berbasis Multi Fungsi Aktivasi Terbobot dalam Prediksi Harga Ethereum", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, hlm. 1101-1107.
10. T. B. S. Alifqi Radjavani. 2023. "Analisa Perbandingan Algoritma CNN Dan LSTM untuk Klasifikasi Pesan Cyberbullying pada Twitter", *Indonesian Journal of Computer Science*, hlm. 1985-1998.