Jurnal Tugas Akhir Fakultas Informatika 12-2020

Implementasi Algoritma Long Short-Term Memory (LSTM) untuk Mendeteksi Penggunaan Kalimat *Abusive* Pada Teks Bahasa Indonesia

Rizka Dwi Wulandari Santosa [1], Moch. Arif Bijaksana [2], Ade Romadhony [3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]rizkadwiwulandari@students.telkomuniversity.ac.id, [2]arifbijaksana@telkomuniversity.ac.id,
[3]aderomadhony@telkomuniversity.ac.id

Abstrak

Penelitian dengan menggunakan Jaringan Syaraf Tiruan atau *Artificial Neural Network* (ANN) sudah banyak dilakukan dan dikembangkan terlebih dalam hal prediksi, klasifikasi dan pendeteksian suatu objek. Salah satu perkembangan dari ANN adalah *Recurrent Neural Network* (RNN). Pada penelitian ini menggunakan salah satu arsitektur dari RNN yaitu Long Short Term Memory (LSTM) yang biasa digunakan untuk masalah deep learning. Arsitektur LSTM diimplementasikan untuk mendeteksi penggunaan kalimat abusive pada teks bahasa indonesia. Dataset yang digunakan pada penelitian mengalami ketidakseimbangan jumlah data pada setiap kelas sehingga dilakukan penambahan data untuk mengetahui pengaruh penambahan jumlah data terhadap hasil performansi arsitektur. Tahapan pengerjaan dalam penelitian ini dimulai dari pembangunan dataset, pra-pemrosesan data, pembuatan model pendeteksi kalimat abusive, pelatihan dan pengujian. Pengujian dilakukan terhadap arsitektur LSTM dan didapatkan hasil bahwa arsitektur ini hanya dapat memprediksi terhadap kelas mayoritas sehingga dilakukan penambahan penggunaan arsitektur yaitu Bidirectional LSTM (BiLSTM). Hasil uji coba menunjukkan BiLSTM lebih baik dalam mengklasifikasikan kalimat karena terdapat forward dan backward layer yang membuat proses pembelajaran model lebih kompleks dalam mengenal konteks kalimat dan hal ini akan meningkatkan keakuratan hasil klasifikasi pada setiap label. Pada LSTM hanya menghasilkan nilai F1 Score untuk kelas mayoritas saja sebesar 0.812 sedangkan pada BiLSTM sudah dapat menghasilkan nilai F1 Score untuk semua kelas.

Kata Kunci: Kalimat Abusive, LSTM, BiLSTM, F1 Score

Abstract

Research using Artificial Neural Network (ANN) has been done and developed especially in terms of prediction, classification and detection of an object. One of the developments of ANN is the Recurrent Neural Network (RNN). In this study, it uses one of the architectures of RNN, Long Short Term Memory (LSTM) which is commonly used for deep learning problems. LSTM architecture is implemented to detect the use of abusive sentences in Indonesian text. The dataset used in the study experienced an imbalance in the amount of data in each class so that the addition of data to find out the effect of increasing the amount of data on the results of architectural performance. The stages of work in this research began from dataset development, data pre-processing, abusive sentence detection modeling, training and testing. Testing was carried out on LSTM architecture and it was obtained that this architecture can only predict against the majority class so that the additional use of architecture is Bidirectional LSTM (BiLSTM). The test results showed that BiLSTM is better at classifying sentences because there are forward and backward layers that make the model learning process more complex in knowing the context of sentences and this will improve the accuracy of classification results on each label. In LSTM only produce F1 Score for majority class only of 0.812 while in BiLSTM can already produce F1 Score for all classes.

Keywords: Kalimat Abusive, LSTM, BiLSTM, F1 Score

1. Pendahuluan

Latar Belakang

Adanya internet dan diikuti dengan munculnya berbagai jenis jejaring sosial merupakan salah satu hasil dari berkembangnya teknologi informasi[1]. Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) melalui buletin APJII November 2020 mengumumkan, jumlah pengguna internet di Indonesia mencapai 196,7 juta pengguna atau setara dengan 73,7% dari total populasi RI yang mencapai 266,9 juta penduduk[2]. Keberadaan internet seharusnya dipergunakan untuk mempermudah pengguna dalam mendapatkan informasi, menjalin hubungan dengan pengguna lain dan memperluas relasi antar pengguna[3]. Namun kemudahan ini tidak serta merta memberikan dampak baik kepada seluruh pengguna internet. Adanya anggapan "ini media sosial saya, terserah saya mau ngomog apa!" kerap kali memicu konflik diantara para pengguna internet. Penggunaan kalimat *abusive* seringkali ditemukan pada unggahan seperti ini dengan tujuan untuk menyerang pihak tertentu atau bahkan hanya sebagai

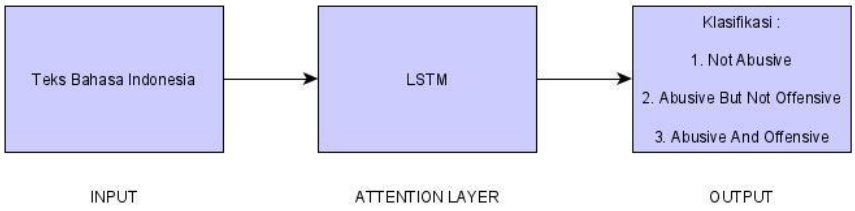
bahan lelucon. Jika kalimat *abusive* dapat dengan mudah ditemukan di internet maka tidak menutup kemungkinan dapat mempengaruhi pola pikir pengguna khusunya yang masih berusia remaja bahwa penggunaan kalimat *abusive* dalam kehidupan sehari-hari tidak masalah[4].

Kalimat *abusive* merupakan ekspresi yang memuat kata-kata kasar atau kotor baik dalam lisan ataupun tulisan. Penyebab banyaknya penggunaan kalimat *abusive* pada internet atau jejaring sosial dikarenakan belum adanya *tools* yang efektif untuk menyaring penggunaan kalimat *abusive*, kurangnya rasa empati antar sesama pengguna internet dan kurangnya pengawasan orang tua[4]. Sebagai contoh, *Hate Speech* atau ujaran kebencian mengandung kalimat *abusive* yang seringkali dapat memicu konflik sosial karena dapat menimbulkan emosi bagi pihak yang dituju maupun siapa saja yang membacanya[5]. Penggunaan kalimat *abusive* juga dapat memberikan dampak buruk bagi kesehatan mental pihak yang dituju terutama jika pihak yang dituju masih tergolong usia remaja[6]. Penggunaan kalimat *abusive* dapat mengarah pada tindakan *cyberbullying* yang mana tingkat depresi yang akan dialami korban dapat lebih tinggi daripada depresi yang didapat dari kekerasan secara fisik[7].

Dalam mengatasi masalah tersebut, dibutuhkan tindakan preventif dan penegakkan hukum yang tegas dan sesuai dengan hukum yang berlaku[8]. Salah satu tindakan preventif bisa dilakukan dengan cara mendeteksi penggunaan kalimat *abusive* untuk penulisan pada media sosial. Deteksi yang dilakukan secara manual akan memakan banyak waktu[9], maka hal ini mendorong peneliti untuk menciptakan cara-cara otomatis yang dapat diterapkan dalam mendeteksi penggunaan kalimat *abusive* pada media sosial. Hal ini dapat dilakukan dengan cara membangun sebuah sistem yang dapat mendeteksi penggunaan kalimat *abusive*. Sistem ini akan dibuat menggunakan salah satu algoritma dari *Recurrent Neural Network* (RNN) yaitu *Long Short Term Memory* (LSTM). Algoritma ini biasa digunakan pada masalah-masalah yang berkaitan dengan *Deep Learning*. Algoritma ini memiliki mekanisme internal yang disebut *gates* atau gerbang yang dapat mengatur aliran informasi. Gerbang ini dapat mempelajari data mana yang penting untuk disimpan atau yang perlu dilupakan dalam sebuah *sequence*[10]. Algoritma LSTM menggunakan *memory cell* yang dapat bekerja lebih baik dibanding dengan jaringan saraf rekuren biasa[11]. Algoritma LSTM juga biasa digunakan untuk prediksi dan klasifikasi[12]. Algoritma LSTM juga cocok digunakan pada data yang memiliki urutan seperti data *time series*, teks dan DNA karena setiap datanya terhubung satu sama lain [13]. Terdapat *imbalanced* atau ketidakseimbangan jumlah antar label pada dataset yang digunakan sehingga terdapat kelas mayoritas dan minoritas. Untuk itu peneliti juga ingin mengetahui pengaruh terhadap adanya penambahan jumlah dataset pada arsitektur yang dibangun.

Topik dan Batasannya

Topik penelitian yaitu mendeteksi penggunaan kalimat *abusive* dengan masukan berupa teks bahasa indonesia lalu diproses dengan algoritma LSTM yang mana hasil keluaran berupa klasifikasi jenis kalimat. Terdapat 3 jenis label dalam klasifikasi hasil keluaran seperti gambar berikut:



Gambar 1. Gambaran Topik Penelitian

Batasan masalah dalam penelitian ini yaitu arsitektur hanya memproses kalimat bahasa indonesia karena *dataset* yang digunakan berupa teks bahasa indonesia sehingga struktur kalimat yang dipelajari hanya struktur kalimat bahasa indonesia. Jumlah *dataset* yang digunakan terdapat imbalanced atau ketidakseimbangan antara jumlah data pada setiap labelnya.

Tujuan

Tujuan dari penelitian ini yaitu mengevaluasi penerapan algoritma LSTM untuk melakukan deteksi penggunaan kalimat *abusive* pada teks bahasa indonesia dan untuk mengetahui pengaruh penambahan data pada dataset kelas minoritas terhadap performansi sistem dalam perhitungan nilai *F1 Score* untuk masing-masing kelas.

Organisasi Tulisan

Pada bagian 2 menjelaskan dasar teori yang digunakan sebagai pedoman dalam penelitian. Pada bagian 3 menjelaskan alur penelitian yang dilakukan. Pada bagian 4 menjelaskan hasil yang didapat melalui algoritma dan sistem yang sudah dibangun. Pada bagian 5 menjelaskan kesimpulan yang didapat dan saran yang bisa dikembangkan untuk penelitian selanjutnya.

2. Studi Terkait

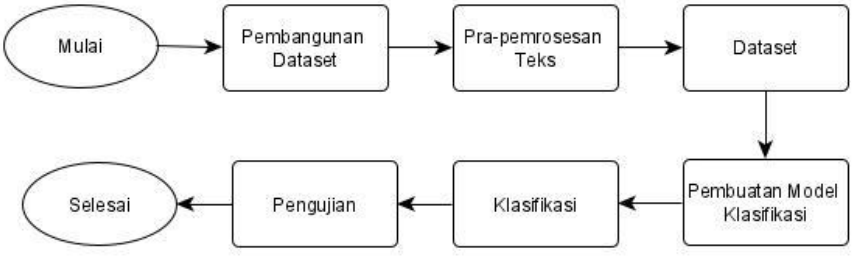
Penelitian ini dibangun berdasarkan beberapa referensi dari penelitian yang sudah dilakukan sebelumnya. Pada penelitian[13], peneliti membangun sebuah sistem untuk mendeteksi kalimat umpatan di media sosial dengan melakukan perbandingan antara dua model *Neural Network* yakni *Artificial Neural Network (ANN)* dan *Recurrent Neural Network (RNN)*. Data yang digunakan berupa *tweet* dengan *track-keywords* tertentu yang sudah didefinisikan sebagai kata yang sering diasosiasikan sebagai umpatan. Data diperoleh dengan menggunakan API Twitter dan diambil secara *stream* dengan pustaka Tweepy pada bahasa pemrograman Phyton dengan jumlah 88,009 data. Pelabelan data berdasarkan preferensi pribadi peneliti dengan membagi menjadi dua label yaitu umpatan (label 1) dan bukan umpatan (label 0). Peneliti juga melakukan teknik *over sampling* dikarenakan jumlah data antar label tidak seimbang dengan menggunakan teknik *synthetic minority over-sampling technique* (SMOTE). Model ANN yang digunakan terdiri dari tiga elemen dasar, yaitu *neuron*s, lapisan (*layers*), dan fungsi aktivasi, serta dua proses utama, yaitu *forward propagate* dan *backward propagate*. Sedangkan model RNN yang digunakan adalah LSTM. Hasil yang didapat dari penelitian menunjukan model RNN menunjukkan performa prediksi kata umpatan atau bukan umpatan yang lebih baik. Dengan pengujian *confusin matrix*, model RNN dapat memprediksi dengan benar 83.8% dari keseluruhan *test set* yang dilabeli sebagai umpatan dan 84.4% yang bukan umpatan sedangkan model ANN hanya dapat memprediksi dengan benar 82.2% yang dilabeli sebagai umpatan dan 80.6% yang bukan umpatan.

Pada penelitian [14], melakukan penelitian tentang mendeteksi ujaran kebencian atau *Hate Speeech* pada kasus Pemilihan Presiden (Pilpres) 2019. *Dataset* yang digunakan diambil dari kolom komentar media sosial Facebook dengan jumlah total 950 kalimat yang mana pada penelitian ini data dibagi menjadi 2 label yaitu *Hate Speech* (HS) dan *Non Hate Speech* (Non_HS). Pada penelitian ini dilakukan 2 kali pengujian dengan jumlah *datatesting* yang berbeda. Penelitian dengan menggunakan LSTM dan Word2Vec ini mendapat nilai akurasi sebesar 58.42%, nilai *recall* sebesar 0.7021, dan nilai *precision* sebesar 0.5641.

Pada penelitian [15], peneliti membuat dataset baru dengan data yang berasal dari media sosial *Twitter* dan didapatkan 2.735 data setelah melalui prapemrosesan teks. Peneliti membuat dua skenario percobaan yaitu pada skenario pertama peneliti mengklasifikasikan tweet menjadi tiga label yakni *bahasa kasar*, *bahasa kasar tapi bukan ujaran kebencian* dan *ujaran kebencian*. Pada skenario kedua peneliti mengklasifikasikan tweet menjadi dua label yaitu *bahasa kasar* dan *bukan bahasa kasar*. Peneliti menggunakan fitur kata *n-gram* dan huruf *n-gram* dengan *Naïve Bayes* (NB), *Support Vector Machine* (SVM) dan *Random Forest Decision Tree* (RFDT) sebagai *classifier*. Hasil penelitian menunjukkan bahwa NB lebih baik dari SVM dan RFDT dengan mencapai nilai 71,15% dari *F1-Score* untuk klasifikasi tiga label dan 87,26 dari *F1-Score* untuk klasifikasi dua label. Hasil penelitian juga menunjukkan pengklasifikasian *tweet* dengan tiga label lebih sulit daripada pengklasifikasian *tweet* dengan dua label.

3. Sistem yang Dibangun

Sistem yang akan dibangun merupakan sistem yang dapat mendeteksi penggunaan kalimat *abusive* pada teks bahasa Indonesia. Gambar 2 merupakan gambaran alur sistem yang akan dibangun pada penelitian ini.



Gambar 2 Alur Pembuatan Sistem

Berdasarkan Gambar 2, alur pembuatan sistem akan dijelaskan sebagai berikut.

3. 1 Pembangunan Dataset

Pada penelitian ini menggunakan dataset yang sudah dibangun dari penelitian[16], dimana data yang digunakan berasal dari komentar berita *online*. Komentar dipilih berdasarkan berita yang sedang *trend* pada bulan maret 2019 hingga september 2019.Total data yang didapatkan sebanyak 3184 komentar. Data terdiri dari tiga label dengan jumlah sebagai berikut.

Table 1 Jumlah Setiap Label Pada Dataset

Label	Keterangan	Jumlah
1	Not Abusive	2779
2	Abusive Not Offensive	100
3	AbusiveAnd Offensive	285

3. 2 Pra-Pemrosesan Teks

Pra-pemrosesan teks adalah tahapan untuk mempersiapkan teks menjadi data yang lebih terstruktur untuk bisa diolah ke tahapan berikutnya. Langkah-langkah yang dilakukan adalah *case folding, removal punctuation, tokenizing, stopword removal, replacing acronym, stemming*.

a. *Case folding* adalah tahapan untuk mengubah semua karakter yang menggunakan huruf kapital menjadi huruf kecil atau upercase.

Tabel 1 Pra-pemrosesan Teks Case Folding

Input	Output
Sebagai ketua KPAI mustinya anda ngurusin masalah anak2 dibawah umur yg ikut demo, tawuran atau yg dilecehkan. Ini kok malah cari2 masalah dng PB Djarum yg jelas2 sdh memberikan banyak sumbangan bagi prestasi bulutangkis nasional. Dasar otak udang loe!!!	sebagai ketua kpai mustinya anda ngurusin masalah anak2 dibawah umur yg ikut demo, tawuran atau yg dilecehkan. ini kok malah cari2 masalah dng pb djarum yg jelas2 sdh memberikan banyak sumbangan bagi prestasi bulutangkis nasional. dasar otak udang loe!!!

b. *Removal Punctuation* adalah tahapan untuk menghilangkan angka,url,mention dan tanda baca yang ada pada kalimat.

Tabel 2 Pra-pemrosesan Teks Removal Punctuation

Input	Output
sebagai ketua kpai mustinya anda ngurusin masalah anak2 dibawah umur yg ikut demo, tawuran atau yg dilecehkan. ini kok malah cari2 masalah dng pb djarum yg jelas2 sdh memberikan banyak sumbangan bagi prestasi bulutangkis nasional. dasar otak udang loe!!!	sebagai ketua kpai mustinya anda ngurusin masalah anak dibawah umur yg ikut demo tawuran atau yg dilecehkan ini kok malah cari masalah dng pb djarum yg jelas sdh memberikan banyak sumbangan bagi prestasi bulutangkis nasional dasar otak udang loe

c. *Replacing acronym* adalah tahap untuk mengganti kata-kata yang disingkat dengan kata aslinya berdasarkan sebuah kamus data singkatan.

Tabel 3 Pra-pemrosesan Teks Replacing Acronym

Input	Output
sebagai ketua kpai mustinya anda ngurusin masalah anak dibawah umur yg ikut demo tawuran atau yg dilecehkan ini kok malah cari masalah dng pb djarum yg jelas sdh memberikan banyak sumbangan bagi prestasi bulutangkis nasional dasar otak udang loe	sebagai ketua kpai mustinya anda ngurusin masalah anak dibawah umur yang ikut demo tawuran atau yang dilecehkan ini kok malah cari masalah dengan pb djarum yang jelas sudah memberikan banyak sumbangan bagi prestasi bulutangkis nasional dasar otak udang loe

d. *Stopword removal* adalah tahapan untuk menghapus kata-kata yang tidak diperlukan seperti kata bantu yang diantaranya adalah 'maka', 'akan', 'yang', 'untuk', 'dan', 'juga', 'dari', 'di' serta 'kan'.

Tabel 4 Pra-pemrosesan Teks Stopword Removal

Input	Output
sebagai ketua kpai mustinya anda ngurusin masalah anak dibawah umur yang ikut demo tawuran atau yang dilecehkan ini kok malah cari masalah dengan pb djarum yang jelas sudah memberikan banyak sumbangan bagi prestasi bulutangkis nasional dasar otak udang loe	sebagai ketua kpai mustinya anda ngurusin masalah anak dibawah umur ikut demo tawuran dilecehkan malah cari masalah dengan pb djarum jelas sudah memberikan banyak sumbangan bagi prestasi bulutangkis nasional dasar otak udang loe

e. *Stemming* adalah tahapan untuk menghilangkan imbuhan pada kata sehingga menjadi kata asli.

Tabel 5 Pra-pemrosesan Teks Stemming

Input	Output
sebagai ketua kpai mustinya anda ngurusin masalah anak dibawah umur ikut demo tawuran dilecehkan malah cari masalah dengan pb djarum jelas sudah memberikan banyak sumbangan bagi prestasi bulutangkis nasional dasar otak udang loe	sebagai ketua kpai musti anda urus masalah anak bawah umur ikut demo tawuran dilecehkan malah cari masalah dengan pb djarum jelas sudah beri banyak sumbangan bagi prestasi bulutangkis nasional dasar otak udang loe

f. *Tokenizing* adalah tahapan untuk memecah kalimat menjadi *list* kata.

Tabel 6 Pra-pemrosesan Teks Tokenizing

Input	Output
sebagai ketua kpai mustinya anda ngurusin masalah anak dibawah umur ikut demo tawuran dilecehkan malah cari masalah dengan pb djarum jelas sudah memberikan banyak sumbangan bagi prestasi bulutangkis nasional dasar otak udang loe	‘sebagai’ ‘ketua’ ‘kpai’ ‘mustinya’ ‘anda’ ‘ngurusin’ ‘masalah’ ‘anak’ ‘dibawah’ ‘umur’ ‘ikut’ ‘demo’ ‘tawuran’ ‘dilecehkan’ ‘malah’ ‘cari’ ‘masalah’ ‘dengan’ ‘pb’ ‘djarum’ ‘jelas’ ‘sudah’ ‘memberikan’ ‘banyak’ ‘sumbangan’ ‘bagi’ ‘prestasi’ ‘bulutangkis’ ‘nasional’ ‘dasar’ ‘otak’ ‘udang’ ‘loe’

3. 3 Dataset

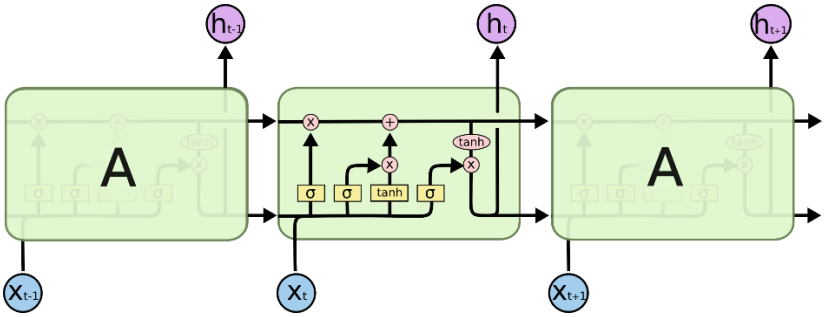
Dataset yang sudah di preprocessing kemudian disimpan dalam *file* ‘dataset-stemming(1)’. Dataset lalu dibagi dengan menggunakan teknik *stratified sampling*. Tujuan dilakukannya teknik *stratified sampling* dikarenakan jumlah dataset yang digunakan pada penelitian ini tidak berimbang untuk setiap kelasnya maka dari itu dataset bisa dikatan *imbalanced*. Dibandingkan dengan metode *simple random sampling*, penggunaan metode *stratified sampling* akan membuat pembagian label lebih seimbang dan lebih informatif untuk kelas minoritas dan kelas mayoritas[17]. Berikut merupakan contoh kalimat pada dataset berdasarkan label yang sudah melalui tahap pra-pemrosesan teks.

Tabel 7 Contoh Kalimat Pada Dataset

Kalimat	Label
setuju emang tidak harus sensor karena anak jadi penasaran malah buat anak cari sendiri	1 (Not Abusive)
goblok maka tidak heran makin banyak orang tidak nonton tv pindah tv cable ytb	2 (Abusive Not Offensive)
kpai goblok cuma pengangguram sudah saat bubar saja tidak penting	3 (Abusive And Offensive)

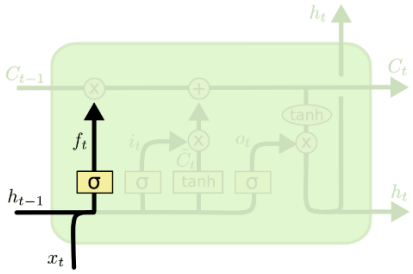
3. 4 Klasifikasi

Pada tahap ini akan dilakukan proses klasifikasi teks bahasa indonesia dengan menggunakan salah satu arsitektur RNN yaitu *Long Short Term Memory* (LSTM). Berikut merupakan penjelasan mengenai arsitektur LSTM[18].



Gambar 3 Arsitektur LSTM

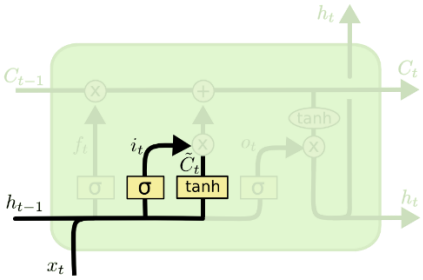
Gambar 3 merupakan arsitektur LSTM dimana pada bagian bawah terdapat *cell gates* yang berfungsi untuk meregulasi informasi yang akan dikeluarkan ke *cell state* atau unit berikutnya. *Cell state* adalah jalur pada bagian atas untuk mengirimkan informasi ke unit selanjutnya.



Gambar 4 Forget Gate Layer

$$ft = \sigma\ (\mathrm{W}f\ [ht\text{-}1\ ,\ xt] + bf) \quad (1)$$

Pada persamaan (1) terdapat formula untuk menghitung keluaran dari lapisan *Forget Gate* dimana informasi akan dihapus dan informasi yang penting akan diteruskan ke *cell state*.

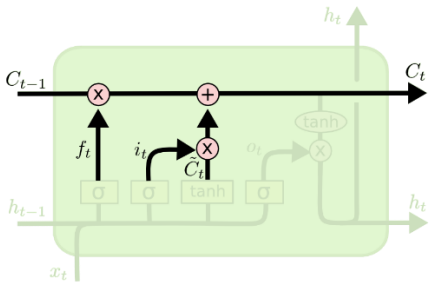


Gambar 5 Input Gate Layer

$$it = \sigma.\ Wi\ [ht\text{-}1\ ,\ xt] + bi) \quad (2)$$

$$\hat{C}t = tanh\ (Wc\ .[ht\text{-}1\ ,\ xt] + bc) \quad (3)$$

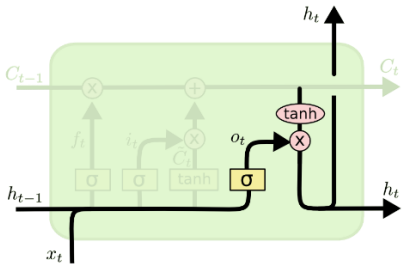
Pada persamaan (2) terdapat formula untuk memasukkan nilai yaitu informasi yang akan diarahkan cell state.



Gambar 6 Update Gate Layer

$$Ct = ft \; x \; Ct\text{-}1 + it \; x \; \hat{C}t \tag{4}$$

Pada persamaan (4) terdapat formula untuk mengubah nilai pada *cell state* yang didapat dari dua lapisan sebelumnya dari proses penghapusan dan penambahan informasi.



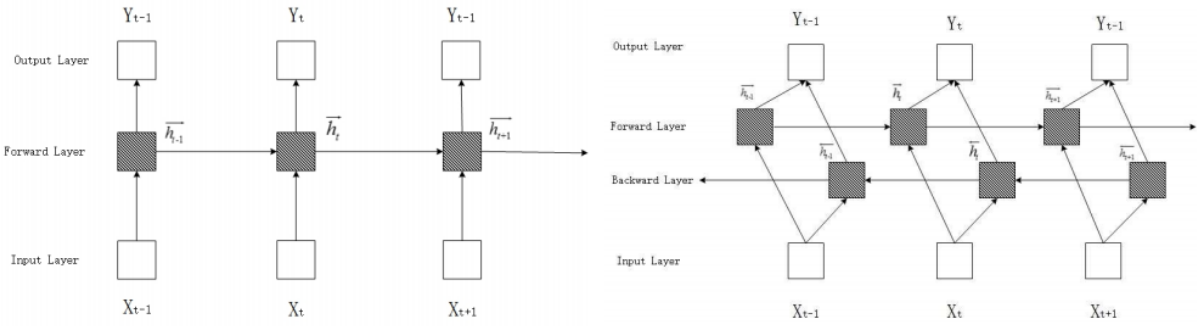
Gambar 7 Output Gate Layer

$$Ot = \sigma \; (Wo \; .[ht\text{-}1 \; , \; xt] + bo) \tag{5}$$

$$ht = Ot \; x \; tanh(Ct \;) \tag{6}$$

Pada persamaan (5) akan menghasilkan hasil yang sesuai untuk diarahkan ke hidden unit berikutnya.

Setelah dilakukan eksperimen hasil dari arsitektur LSTM menunjukkan performansi yang kurang baik terhadap kelas minoritas sehingga dilakukan penambahan eksperimen pada jenis arsitektur yang digunakan yaitu *Bidirectional Long Short Term Memory* (BiLSTM). Pada dasarnya BiLSTM memiliki struktur yang sama dengan LSTM dengan adanya penambahan *backward layer*. Ilustrasi dari arsitektur LSTM dan BiLSTM sebagai berikut.



(a) LSTM (b) BiLSTM

Gambar 8 Ilustrasi Arsitektur LSTM dan BiLSTM

BiLSTM merupakan pengembangan dari arsitektur LSTM dimana terdapat dua lapisan yang prosesnya saling berlawanan arah dimana arsitektur ini cocok untuk digunakan dalam mengenali pola dalam kalimat karena setiap kata diproses secara sekuensial. *Forward layer* bergerak maju atau ke kanan dimana lapisan ini memahami dan memproses dari kata pertama menuju kata terakhir sedangkan *backward layer* bergerak mundur atau ke kiri dimana

lapisan ini memahami dan memproses dari kata terakhir menuju kata pertama. Berikut merupakan nilai *hyperparameters* yang diterapkan pada arsitektur LSTM dan BiLSTM pada penelitian ini.

Tabel 8 Nilai Hyperparameters Arsitektur LSTM dan BiLSTM

Parameter	Nilai
Vocab_size	6000
Embedding_dim	64
Dense	3
Fungsi Aktivasi	*Softmax*
Learning rate	0,001
Fungsi Loss	*Sparse categorial crossentrophy*
Jumlah epoch	100 (*Early Stopping*)

Penggunaan *Early Stopping* digunakan agar model yang sudah dibangun berhenti melakukan *training* pada saat nilai validasi *error* atau validasi *loss* mencapai nilai minimum [19] atau pada saat nilai validasi loss tidak mengalami improvisasi.

3. 5 Evaluasi Sistem

Evaluasi sistem dilakukan dengan menggunakan *confusion matrix multiclass* dan menghitung nilai akurasi, *precision, recall dan F1-score* dari setiap kelas. *Confusion matrix multiclass* adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan. Tabel *confusion matrix multiclass* yang diterapkan pada penelitian ini adalah sebagai berikut.

Tabel 9 Confution Matrix Multiclass

		Predicted		
Actual		**Not Abusive**	**Abusive Not Offensive**	**Abusive And Offensive**
	Not Abusive	TP NotAbusive	X	X
	Abusive Not Offensive	X	TP Abusive Not Offensive	X
	Abusive and Offensive	X	X	TP Abusive And Offensive

Pada tabel 10 terdapat tiga label prediksi sesuai dengan dataset. TP adalah singkatan dari *True Positive* yang merupakan kasus dimana prediksi dan nilai yang sebenarnya bernilai *True* atau benar. Pada *confusion multiclass matrix* hanya tertera TP karena untuk penentuan FN *(False Negative)* berasal dari seluruh jumlah baris per-label sedangkan untuk penentuan FP *(False Positive)* berasal dari seluruh jumlah kolom per-label dan *TN (True Negative)* adalah saat nilai prediksi tidak ada dan nilai aktualnya salah[20]. Performa dari deteksi penggunaan kalimat *abusive* diukur dengan menggunakan beberapa parameter yaitu *precision, recall* dan *F1 Score*. Formula dari parameter-parameter tersebut diberikan pada persamaan (7-9).

Precision adalah perbandingan antara data yang diklasifikasikan secara benar dibandingan dengan seluruh data yang diklasifikasikan secara benar.

$$Precision = \frac{TP}{(TP+FP)} \quad (7)$$

Recall adalah perbandingan antara data yang diklasifikasikan secara benar dengan jumlah data yang berada di kelas tersebut. Rumus untuk masing-masing perhitungan adalah sebagai berikut.

$$Recall = \frac{TP}{(TP+FN)} \quad (8)$$

F1 Score adalah perbandingan rata-rata nilai *precision* dan nilai *recall* yang dibobotkan.

$$F1\ Score = 2 \text{ x } \frac{precision\ x\ recall}{precision + recall} \quad (9)$$

4. Evaluasi dan Analisis

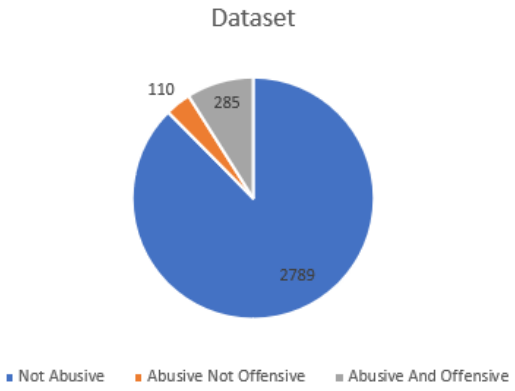
Eksperimen pertama dilakukan dengan membandingkan hasil *F1 Score* dari semua kelas menggunakan metode *non-neural network* yaitu K-Nearest Neighbors (KNN), Naive Bayes dan SVM dengan metode RNN yaitu LSTM dan BiLSTM. Hasil perbandingan nilai F1 Score yang didapatkan dari kedua arsiterktur seperti pada tabel 11.

Tabel 10 Hasil F1 Score KNN dan LSTM

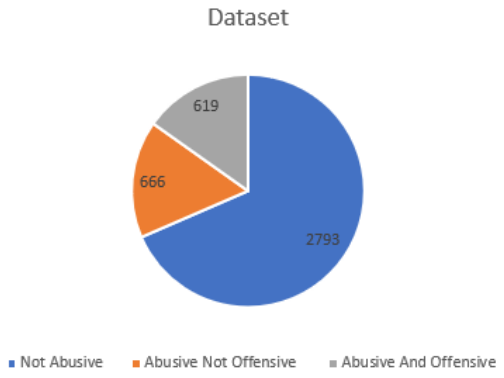
Arsitektur	**Jenis Arsitektur**	***F1 Score* kelas 1**	***F1 Score* Kelas 2**	***F1 Score* Kelas 3**
KNN	Non Neural	0.9331	0	0
Naive Bayes	Non Neural	0.3319	0.0863	0.2368
SVM	Non Neural	0.9339	0	0
LSTM	Recurrent Neural	0.9331	0	0
BiLSTM	Recurrent Neural	0.9375	0	0.2353

Dari hasil eksperimen pertama menunjukkan bahwa arsitektur KNN, SVM dan LSTM hanya dapat mengklasifikasikan data terhadap kelas mayoritas atau kelas Not Abusive dan kedua arsitektur belum mampu mengklasifikasikan data terhadap kelas minoritas. Dari hasil *F1 Score* pada tabel 11 juga menunjukkan bahwa arsitektur BiLSTM yang merupakan salah satu dari arsitektur RNN lebih baik dalam mengklasifikasikan kalimat daripada arsitektur KNN, Naive Bayes dan SVM yang merupakan arsitektur *non-neural network*.

Untuk mencapai tujuan penelitian dan mengetahui performansi dari arsitektur maka dilakukan pengujian skenario dalam mengetahui pengaruh penambahan jumlah dataset terhadap arsitektur LSTM dan arsitektur BiLSTM yang sudah dibangun. Pada skenario pengujian ini dilakukan pengujian dengan menggunakan dua dataset dengan jumlah yang berbeda. Pada pengujian pertama dataset berjumlah 3.184 kalimat dimana terdiri dari 2.789 kalimat berlabel 1 atau Not Abusive, 110 kalimat berlabel 2 atau Abusive Not Offensive dan 285 kalimat berlabel 3 atau Abusive And Offensive. Pada pengujian yang kedua, terdapat penambahan data pada dataset yang berasal dari penelitian [5]. Dataset tersebut dipilih karena dataset berisikan kalimat dalam bahasa indonesia dengan pelabelan pada kalimatnya memiliki kesamaan makna dengan dataset yang sudah digunakan pada awal penelitian ini. Penambahan dataset sebanyak 894 kalimat dimana terdiri dari 2.793 kalimat berlabel 1 atau Not Abusive, 666 kalimat berlabel 2 atau Abusive Not Offensive dan 619 Kalimat berlabel 3 atau Abusive And Offensive yang mana jumlah total data menjadi 4.078 kalimat. Jumlah data dalam dataset pengujian pertama dapat dilihat pada gambar 9 dan dataset pengujian kedua dapat dilihat pada gambar 10.



Gambar 9 Jumlah Kalimat Setiap Label Dataset Pertama



Gambar 10 Jumlah Kalimat Setiap Label Dataset Kedua

Dari pengujian degan menggunakan jumlah data yang berbeda didapatkan hasil F1 Score pada setiap kelasnya seperti berikut.

Tabel 11 Hasil Pengujian Skenario Pengaruh Penambahan Dataset

Arsitektur	*F1 Score* kelas 1		*F1 Score* Kelas 2		*F1 Score* Kelas 3		*Weighted Avg F1 Score*	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2
LSTM	0.9331	0.8128	0	0	0	0	0.8181	0.5565
BiLSTM	0.9375	0.9097	0	0.6293	0.2353	0.5089	0.8423	0.8030

Berdasarkan tabel 12, pada pengujian dengan menggunakan dataset pertama arsitektur LSTM hanya mampu mengklasifikasikan data berdasarkan data pada kelas 1 (Not Abusive) atau kelas mayoritas saja sedangkan pada arsitektur BiLSTM sudah dapat mengklasifikasikan data pada kelas 1 (Not Abusive) dan kelas 3 (Abusive and Offensive) yang mana terjadi peningkatan yaitu arsitektur BiLSTM sudah mampu mengklasifikasikan salah satu kelas minoritas. Setelah dilakukan penambahan jumlah data pada dataset kedua, arsitektur LSTM masih belum mampu mengklasifikasikan data pada kelas minoritas sedangkan pada arsitektur BiLSTM sudah mampu mengklasifikasikan data pada semua kelas dan menghasilkan nilai *F1 Score* untuk semua kelas dari dataset.

Berdasarkan pengujian arsitektur terhadap data tes yang sudah dilakukan, dilakukan analisis pada kalimat-kalimat yang belum sesuai diklasifikasikan atau belum sesuai dideteksi oleh arsitektur dan didapatkan beberapa alasan sebagai berikut:

1. Terdapat beberapa kata kasar yang berubah maknanya setelah dilakukan pra-pemrosesan teks. Contoh kata ‘bajingan’ setelah dilakukan pra-pemrosesan teks dan melalui tahap *stemming* maka menjadi kata ‘bajing’.
2. Terdapat beberapa kalimat yang penulisan kata kasarnya dengan menggabungkan huruf dan angka sehingga pada saat dilakukan pra-pemrosesan teks maka penulisan dari kata kasar tersebut berubah. Contoh kata ‘Tol0l’ menjadi ‘toll’.
3. Terdapat beberapa penulisan kata kasar dengan mengganti beberapa hurufnya. Contoh kata ‘bangsad’, ‘geblek’, ‘anying’.
4. Terdapat penggunaan kata kasar yang disingkat atau dihilangkan huruf vokalnya sehingga arsitektur tidak mengenali kesamaan makna kata tersebut dengan kata kasar yang serupa. Contoh kata ’gblg’.
5. Terdapat penggunaan kata kasar dalam bahasa asing dan bahasa daerah yangmana membuat arsitektur tidak mengenali kata tersebut sebagai kata kasar yang sudah dipelajari. Contoh kata ‘ndasmu’, ‘stupid’.
6. Terdapat beberapa kalimat yang ditujukan untuk menghina namun tidak menggunakan kata-kata kasar atau hinaan yangmana arsitektur belum mampu untuk mengenali konteks dengan kalimat hinaan yang dituliskan secara implisit. Contoh kalimat seperti ‘si zonk kalo baca komen di detik com mesti malu kalo perlu undur diri tapi sayang dia sdh gak punya malu muka tembok mulut nyinyir adalah takdir’.

5. Kesimpulan

Kesimpulan yang didapatkan dari penelitian ini adalah sebagai berikut.

1. Arsitektur LSTM belum mampu mendeteksi kalimat *abusive* dengan optimal terutama pada kasus dataset yang tidak berimbang jumlahnya atau mengalami *imbalanced*. Pada pengujian dengan dataset pertama, LSTM hanya dapat memprediksi data ke dalam kelas mayoritas, sedangkan pada BiLSTM sudah dapat mengklasifikasikan data ke kelas nmayoritas atau kelas 1 dan salah satu kelas minoritas yaitu kelas 2.
2. Adanya penambahan data pada kelas minoritas pada dataset kedua, arsitektur LSTM masih belum mampu mengklasifikasikan data kelas minoritas dan hanya menghasikan nilai *F1 Score* dari kelas mayoritas. Hal ini dikarenakan pada arsitektur LSTM hanya terdapat *forward layer* yang mana *layer* ini bergerak dari kiri ke kanan atau dari kata pertama menuju kata terakhir dalam proses pembelajaran model dalam mengenal konteks kalimat.
3. Pada BiLSTM sudah dapat mengklasifikasikan ketiga jenis kelas dan menghasilkan semua nilai *F1 Score* untuk setiap kelas. Hal ini dikarenakan pada BiLSTM terdapat *forward layer dan backward layer* yang mana pada *backward layer* bergerak dari kanan ke kiri atau dari kata terakhir menuju kata pertama sehingga proses pembelajaran model akan lebih kompleks dalam mengenal konteks kalimat yang mana hal ini akan meningkatkan keakuratan hasil klasifikasi pada setiap label.
4. Penambahan jumlah dataset juga berpengaruh pada hasil *F1 Score* dari arsitektur BiLSTM. Hasil *F1 Score* arsitektur BiLSTM mengalami peningkatan setelah dilakukan adanya penambahan data pada kelas minoritas.

Reference

[1] Y. Fitriani, "Analisis Pemanfaatan Berbagai Media Sosial sebagai Sarana Penyebaran Informasi bagi Masyarakat," *Paradig. - J. Komput. dan Inform.*, vol. 19, no. 2, pp. 148–152, 2017, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/paradigma/article/view/2120.

[2] Asosiasi Penyelenggara Jasa Internet Indonesia, "BULETINAPJIIEDISI74November2020.pdf."

[3] B. A. Simangunsong, "Interaksi Antarmanusia melalui Media Sosial Facebook Mengenai Topik Keagamaan," *J. ASPIKOM*, vol. 3, no. 1, p. 65, 2016, doi: 10.24329/aspikom.v3i1.99.

[4] S. Tuarob and J. L. Mitrpanont, "Automatic Discovery of Abusive Thai Language Usages in Social Networks," in *Digital Libraries: Data, Information, and Knowledge for Digital Lives*, 2017, pp. 267–278.

[5] M. O. Ibrohim and I. Budi, "Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter," pp. 46–57, 2019, doi: 10.18653/v1/w19-3506.

[6] Z. Xu and S. Zhu, "Filtering offensive language in online communities using grammatical relations," *7th Annu. Collab. Electron. Messag. Anti-Abuse Spam Conf. CEAS 2010*, 2010.

[7] Rahmat Syah; Istiana Hermawati, "Upaya pencegahan kasus cyberbullying bagi remaja pengguna media sosial di indonesia," *J. Penelit. Kesejaht. Sos.*, vol. 17 no 2, no. 2, pp. 131–146, 2018, [Online]. Available: https://www.elearningkebencanaan.education/longsor/upaya-pencegahan-longsor/.

[8] P. E. Septiani, "Jurnal Pengabdian Masyarakat," *Din. J. Pengabdi. Kpd. Masy.*, vol. 3, no. 1, pp. 105–111, 2019, doi: 10.31849/dinamisia.v3i1.2729.

[9] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *15th Conf. Eur. Chapter Assoc. Comput. Linguist. EACL 2017 - Proc. Conf.*, vol. 2, no. 2, pp. 427–431, 2017, doi: 10.18653/v1/e17-2068.

[10] D. Pratidana, "Hak cipta dan penggunaan kembali : Lisensi ini mengizinkan setiap orang untuk menggubah , memperbaiki , dan membuat ciptaan turunan bukan untuk kepentingan komersial , selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat ya," *J. Exp. Psychol. Gen.*, vol. 136, no. 1, pp. 23–42, 2017, [Online]. Available: http://kc.umn.ac.id/5548/1/BAB II.pdf.

[11] M. Roondiwala, H. Patel, and S. Varma, "Predicting Stock Prices Using LSTM," *Int. J. Sci. Res.*, vol. 6, no. 4, pp. 2319–7064, 2015, [Online]. Available: https://www.quandl.com/data/NSE.

[12] N. K. Manaswi, "RNN and LSTM," in *Deep Learning with Applications Using Python : Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras*, Berkeley, CA: Apress, 2018, pp. 115–126.

[13] S. Sahrul, A. F. Rahman, M. D. Normansyah, and A. Irawan, "Sistem Pendeteksi Kalimat Umpatan Di Media Sosial Dengan Model Neural Network," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 3, no. 2, pp. 108–115, 2019.

[14] A. S. Talita and A. Wiguna, "Implementasi Algoritma Long Short-Term Memory (LSTM) Untuk Mendeteksi Ujaran Kebencian (Hate Speech) Pada Kasus Pilpres 2019," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 37–44, 2019, doi: 10.30812/matrik.v19i1.495.

[15] lucia maria aversa Villela, "PENDETEKSIAN BAHASA KASAR (ABUSIVE LANGUAGE) DAN UJARAN KEBENCIAN (HATE SPEECH) DARI KOMENTAR DI JEJARING SOSIAL," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.

[16] D. R. K. Desrul and A. Romadhony, "Abusive Language Detection on Indonesian Online News Comments," *2019 2nd Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2019*, pp. 320–325, 2019, doi: 10.1109/ISRITI48646.2019.9034620.

[17] Q. Wu, Y. Ye, H. Zhang, M. K. Ng, and S. S. Ho, "ForesTexter: An efficient random forest algorithm for imbalanced text categorization," *Knowledge-Based Syst.*, vol. 67, pp. 105–116, 2014, doi: 10.1016/j.knosys.2014.06.004.

[18] D. J. M. Pasaribu, K. Kusrini, and S. Sudarmawan, "Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon dengan Bidirectional LSTM dan Bert Embedding," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, pp. 9–20, 2020, doi: 10.35585/inspir.v10i1.2568.

[19] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.

[20] V. W. Siburian and I. E. Mulyana, "Prediksi Harga Ponsel Menggunakan Metode Random Forest," *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 144–147, 2018.