Terbit online pada laman web jurnal: **https://ejurnalunsam.id/index.php/jicom/**

Jurnal Informatika dan Teknologi Komputer
Vol. 05 No. 02 (2024) 126 - 132 E-ISSN :2774-7115 P-ISSN: 2775-2089

Analisis Sentimen Ulasan Film Menggunakan Metode BiLSTM

Muhammad Dzaki Arkaan Nasir[1], Syarif Hidayat[2]
[1]Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia
[2]Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia
[1]19523212@students.uii.ac.id, [2]syarif@uii.ac.id*

Abstract

In the digital age, film reviews on social media platforms and review sites have become an important source of information about public opinion. However, manually processing a large number of film reviews is very difficult and time-consuming. Therefore, this research focuses on sentiment analysis of film reviews using natural language processing (NLP) techniques and BiLSTM, an extension of recurrent neural network (RNN) models. Using a data set of 50,000 film reviews from IMDB, this study ran four different scenarios to train the BiLSTM model. Results showed that Model 3, with Embedding Dimension 64 and Batch Size 32, gave the best performance with an accuracy of 0.8727 which is a good accuracy as it means that 43,635 out of 50,000 reviews were predicted correctly. The research conclusion highlights the effectiveness of BiLSTM in analyzing sentiment on movie reviews, particularly in overcoming the challenges of long text.

Keywords: sentiment analysis; natural language processing (NLP); movie reviews, BiLSTM, IMDb

Abstrak

Dalam era digital, ulasan film di platform media sosial dan situs ulasan menjadi sumber penting informasi tentang opini publik. Namun, mengolah sejumlah besar ulasan film secara manual sangatlah sulit dan memakan waktu. Oleh karena itu, penelitian ini fokus kepada analisis sentimen ulasan film dengan menggunakan teknik pemrosesan bahasa alami (NLP) dan BiLSTM, sebuah perluasan dari model jaringan saraf berulang (RNN). Menggunakan *data set* yang terdiri dari 50.000 ulasan film dari IMDb, penelitian ini menjalankan empat skenario berbeda untuk melatih model BiLSTM. Hasil menunjukkan bahwa Model 3, dengan *Embedding Dimension* 64 dan *Batch Size* 32, memberikan kinerja terbaik dengan akurasi 0,8727 hal ini merupakan akurasi yang baik karena dengan kata lain sebanyak 43.635 dari 50.000 ulasan diprediksi dengan benar. Kesimpulan penelitian menyoroti keefektifan BiLSTM dalam menganalisis sentimen pada ulasan film, khususnya dalam mengatasi tantangan teks panjang.

Kata kunci: analisis sentimen, pemrosesan bahasa alami (NLP), ulasan film, BiLSTM, IMDb.

1. Pendahuluan

Dalam beberapa tahun terakhir, perkembangan teknologi telah mengubah bagaimana cara manusia berinteraksi dengan dunia digital secara signifikan. Situs ulasan film seperti IMDb dan Rotten Tomatoes yang dimana terdapat ratusan ribu ulasan film telah diposting setiap harinya di platform-platform tersebut telah menjadi tempat populer bagi individu. Ulasan tersebut bukan hanya merepresentasikan pendapat, namun juga pengalaman penonton dan sudut pandang mereka mengenai film-film tertentu. Dengan volume ulasan yang semakin bertambah setiap harinya, pemahaman terhadap sentimen yang terkandung dalam ulasan tersebut memiliki nilai penting dalam mengukur reaksi dan tanggapan penonton terhadap suatu karya film tertentu.

Pendekatan manual dalam mengolah ulasan sudah tidak lagi praktis dan memakan banyak waktu dengan volume data yang sebesar itu. Oleh karena itu, penelitian sebelumnya mulai berfokus menggunakan metode analisis sentimen. Dalam konteks ini, analisis sentimen merupakan metode yang menggabungkan pemrosesan bahasa alami/*natural language processing* (NLP) dan *machine learning* (ML) untuk mengekstrak, mengidentifikasi serta mengklasifikasikan berbagai pendapat yang telah diungkapkan. Sejalan dengan yang diungkapkan pada penelitian [1] metode ini memungkinkan kita untuk memahami pandangan dan emosi orang dalam suatu ulasan. Analisis sentimen pada dasarnya adalah masalah klasifikasi yang berada di antara NLP dan ML [2].

Analisis sentimen memungkinkan kita untuk menganalisis dan memahami opini, pandangan, dan perasaan orang terhadap subjek, termasuk ulasan film serta mengidentifikasi dan mengukur tingkat sentimen positif dan negatif dalam ulasan-ulasan. Analisis

sentimen terhadap ulasan film menjadi sangat berguna karena dapat digunakan untuk memberikan wawasan bagi pembuat film, studio produksi, dan penonton lain yang ingin menilai kualitas sebuah film sebelum menontonnya.

Pada masa lampau terdapat tantangan yang dihadapi. Salah satunya menunjukkan bahwa analisis sentimen tradisional seringkali kesulitan dalam mengatasi nuansa bahasa dan konteks dalam suatu film [3]. Namun, masalah tersebut dapat teratasi dengan perkembangan dalam NLP dan teknologi jaringan syaraf, khususnya Recurrent Neural Network (RNN).

RNN dengan kemampuannya untuk memanfaatkan informasi kontekstual dari data sekuensial, telah terbukti berharga dalam analisis sentimen [4]. Namun, model RNN memiliki beberapa keterbatasan, seperti kesulitan mengatasi masalah vanishing gradient dan exploding gradient, yang menghambat kemampuannya dalam menangkap ketergantungan jangka panjang [5]. Akibatnya, performa model RNN sering kali menurun ketika dihadapkan pada tugas analisis sentimen pada teks panjang, seperti ulasan film yang bisa mengandung banyak kalimat [6].

Dalam upaya untuk mengatasi kendala dalam memahami konteks urutan pada data teks panjang, para peneliti telah mengembangkan berbagai jenis model jaringan saraf yang lebih canggih. Dua di antaranya adalah LSTM (Long Short-Term Memory) dan BiLSTM (Bidirectional LSTM).

LSTM merupakan pengembangan model dari RNN yang dirancang untuk mengatasi masalah vanishing gradient yang biasa terjadi pada RNN standar [7]. Penelitian yang dilakukan Gers dkk.[8] berkontribusi dengan menciptakan sel memori yang bisa mempertahankan informasi dalam waktu yang lama. Selain itu, mereka juga memperkenalkan gerbang gerbang input dan gerbang output serta gerbang forget yang membuat LSTM menjadi lebih efisien dalam memahami ketergantungan jangka panjang dalam data sekuensial. Gerbang forget juga memungkinkan mengontrol banyaknya informasi yang tidak lagi diperlukan dari sel memori [9].

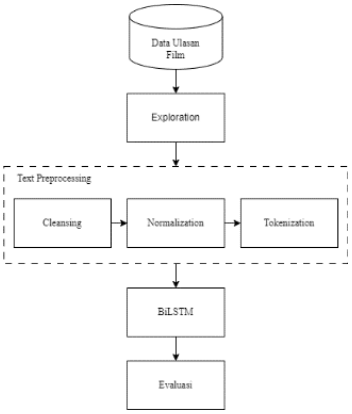
BiLSTM adalah pengembangan lebih lanjut dari model LSTM, di mana informasi tidak hanya dipelajari dari masa lalu ke masa depan, tetapi juga sebaliknya, dari masa depan ke masa lalu. Sebuah jaringan BiLSTM sebenarnya terdiri dari dua LSTM yang terpisah: satu belajar melalui urutan input seperti biasa (dari awal ke akhir), sedangkan yang lain belajar melalui urutan yang dibalik (dari akhir ke awal) [10]. Sehingga BiLSTM dapat menangkap konteks lebih baik daripada LSTM [11].

Kedua model ini telah terbukti efektif dalam berbagai tugas NLP, termasuk analisis sentimen pada teks panjang seperti ulasan film. Hal ini juga dibuktikan pada studi [12] menunjukkan bahwa BiLSTM memiliki akurasi sebesar 86% dan LSTM memiliki akurasi 85% dalam melakukan analisis sentimen pengguna Twitter terhadap layanan rumah sakit saat pandemi Covid-19. Oleh karena itu, dalam penelitian ini, kami akan membandingkan kinerja LSTM dan BiLSTM dalam analisis sentimen pada ulasan film, dengan harapan dapat menemukan metode yang lebih unggul untuk tugas tersebut.

2. Metode Penelitian

Penelitian ini menggunakan *data set* yang digunakan untuk penelitian sebelumnya oleh (Mass dkk., 2011) yang terdiri dari 50000 ulasan film dengan menggunakan metode BiLSTM dengan tujuan penelitian untuk menilai seberapa besar akurasi metode BiLSTM dalam menganalisis sentimen dari ulasan film.

2.1. Alur Pelaksanaan Penelitian



Gambar 1. Alur Penelitian

Penelitian ini memiliki alur yang meliputi pengumpulan *data set*, *exploration*, *preprocessing text*, pembuatan model BiLSTM, dan evaluasi akurasi.

2.2 Pengumpulan Data set

Data set ulasan film ini berasal dari penelitian [13] yang terdiri dari 50.000 ulasan film dari situs IMDb yang dibatasi tidak lebih dari 30 ulasan per filmnya dengan komposisi ulasan positif dan negatif yang seimbang dengan ulasan positif memiliki skor lebih ≥ 7 dari 10 dan ulasan negatif ≤ 4 dari 10. Ulasan netral tidak termasuk dalam *data set* ini.

2.3 Exploration

Exploration atau *Exploratory Data Analysis* (EDA) dalam NLP merupakan proses awal untuk memahami karakter struktur, pola, anomali, dan karakteristik lainnya dari data teks yang dimiliki. Tujuan dari dilakukannya EDA adalah untukmendapatkan wawasan yang mendalam tentang data sebelum memproses atau membangun model.

2.4 Preprocessing Text

Prepocessing text adalah proses merubah teks yang tidak terstruktur menjadi lebih terstruktur. Teks juga dirubah menjadi bentuk numerik [14]. Beberapa tahapan pada *preprocessing text* antara lain:

1. Cleansing
   Tahap ini merupakan proses eliminasi bagian teks yang tidak penting. Proses yang dilakukan meliputi penghapusan karakter seperti tanda baca, angka, karakter spesial, dan *stopword*.
2. Normalization
   Pada proses *normalization* teks diubah menjadi bentuk standar, seperti merubah menjadi huruf kecil dan proses *lemmatization* atau mengubah kata menjadi bentuk dasar.
3. Tokenization & Indexing
   Pada tahap *tokenization*, setiap kalimat akan dipish menjadi unit yang lebih kecil dalam bentuk per kata. Setiap kata yang sudah dipecah disimpan ke dalam unit atau disebut sebagai *"Token"* [15].

2.5 Pembuatan Model BiLSTM

Pada tahap ini, model akan dibangun dengan arsitektur BiLSTM yang terdiri dari beberapa lapisan/*layer* seperti:

1. Embedding Layer
   Layer ini akan mengkonversi kata dalam teks menjadi representasi vektor numerik. Ini memungkinkan model dapat memahami hubungan semantik antar kata [4].
2. BiLSTM Layer
   Lapisan BiLSTM terdiri dari dua lapisan LSTM yang dijalankan secara pararel dengan arah maju (*forward)* dan arah mundur (*backward).* Tahap ini memungkinan model menggabungkan informasi dari kedua arah dalam urutan data [16].
3. Dropout Layer
   Lapisan *dropout* digunakan untuk mencegah *overfitting* dalam model. Neuron dipilih secara acak dan dihilangkan selama pelatihan. Ini memaksa jaringan untuk mempelajari representasi yang lebih kuat dan tidak bergantung pada neuron tertentu
4. Dense Layer
   Dense Layer atau biasa dikenal sebagai *fully connected layer*, menghubungkan setiap neuron di layer sebelumnya dengan setiap neuron di layer berikutnya.

2.6 Evaluasi Akurasi

Tahap evaluasi ini akan menggunakan Confusion Matrix. Confusion matrix, juga dikenal sebagai error matrix, adalah alat visualisasi yang sering digunakan dalam machine learning dan statistik untuk memahami kinerja model klasifikasi. Matriks ini memberikan empat kombinasi berbeda dari nilai prediksi dan nilai aktual yang dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

	Predicted Negative	*Predicted Positive*
Actual Negative	*True Negative (TN)*	*False Positive (FP)*
Actual Negative	*False Negative (FN)*	*True Positive (TP)*

Penjelasan dari tabel di atas adalah nilai *True Positive* (TP) adalah nilai yang diprediksi dengan nilai aktualnya memiliki nilai positif. *True Negative* (TN) adalah nilai yang diprediksi dengan nilai aktualnya memiliki nilai negatif. *False Positive* (FP) ketika nilai yang diprediksi bernilai positif tetapi aktualnya bernilai negatif. *False Negative* (FN) ketika nilai yang diprediksi bernilai negatif dan aktualnya memiliki nilai positif [17] Ini memungkinkan untuk menghitung *Accuracy*, *precision*, *recall*, dan *F1-score*.

Accuracy yaitu menghitung seberapa akurat model memprediksi benar atau nilai prediksi sama dengan nilai aktual dibandingkan dengan keseluruhan nilai. Accuracy dapat dihitung menggunakan persamaan 1.

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

Precision mengukur sejauh mana model mampu meminimalkan jumlah kesalahan dalam mengklasifikasikan objek sebagai positif ketika sebenarnya objek tersebut negatif. Precision dapat dihitung dengan menggunakan persamaan 2.

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

Recall mengukur kemampuan model dalam mengidentifikasi objek aktual positif dari semua objek yang sebenarnya positif (TP dan TN). *Recall* dapat dihitung dengan persamaan 3.

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

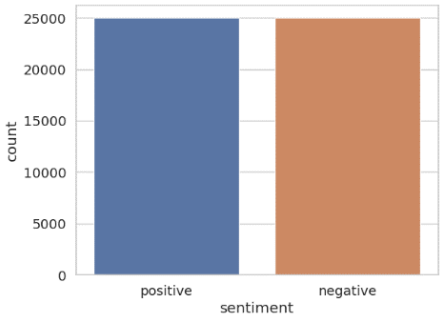
F1-Score adalah metrik yang menggabungkan *precision* dan *recall* menjadi satu metrik tunggal. *F1-Score* berguna ketika data tidak memiliki keseimbangan antar klasifikasi. *F1-Score* dapat dihitung dengan menggunakan persamaan 4.

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

3. Hasil dan Pembahasan

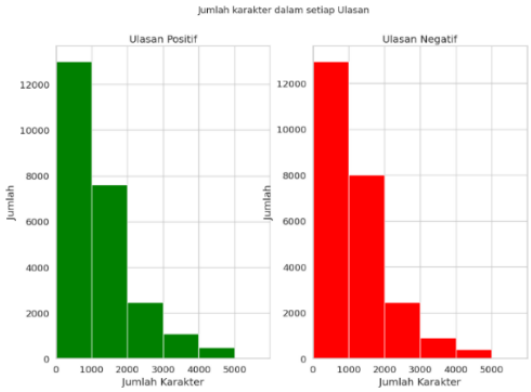
3.1. Exploration

Analisis awal terhadap data teks di fokuskan terhadap sebaran sentimen dalam *data set*. Menurut Gambar 2, *data set* ini memiliki distribusi yang seimbang antara sentimen positif dan sentimen negatif, yang merupakan kondisi ideal untuk pemodelan. Keseimbangan ini sangat penting karena dapat membantu pengembangan model dan menghindari *overfitting.*



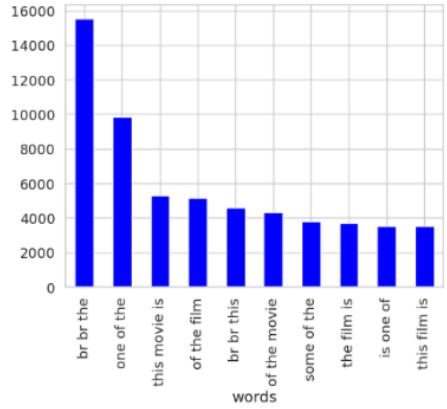
Gambar 2. Sebaran Sentimen Dalam *Data set*

Kemudian dilakukan analisis jumlah karakter tiap ulasan yang di tampilkan pada Gambar 3. Sebagian besar ulasan baik ulasan positif maupun ulasan negatif berada dalam rentang 0-1000 karakter. Namun, terdapat perbedaan dalam jumlah karakter antara ulasan positif dan ulasan negatif khususnya dalam rentang 1000-2000 karekter. Ulasan negatif cenderung lebih panjang dalam rentang tersebut, hal tersebut menunjukkan bahwa penonton mungkin mengekspresikan kekecewaan atau kritik mereka dengan lebih detail dibandingkan dengan ulasan positif. Ini bisa mencerminkan kecenderungan untuk memberikan penjelasan yang lebih panjang ketika mereka tidak puas dengan sebuah film.



Gambar 3. Jumlah Karakter Tiap Sentimen Sebelum *Preprocessing text*

Analisis dilanjutkan kepada analisis trigram atau kombinasi tiga kata berurutan yang diambil dari teks. Pada Gambar 4 dapat dilihat sepuluh trigram dengan frekuensi kemunculan terbanyak dalam *data set*.



Gambar 4. Sepuluh Trigram Dengan Frekuensi Terbanyak Sebelum *Preprocessing text*

Ketika memperhatikan Gambar 4, dapat dilihat Terdapat beberapa elemen yang tidak membawa makna, seperti "br" yang merupakan tag HTML dan stopwords seperti "of", "the", dan "is". Keberadaan kata-kata ini tidak menambah nilai dalam analisis sentimen dan hanya akan memberatkan model nantinya.

3.2 Preprocessing text

Berdasarkan dari temuan sebelumnya, dilakukan proses preprocessing text. Dengan menghilangkan kata-kata yang tidak memiliki makna, kita bisa meningkatkan akurasi model dalam mengklasifikasikan sentimen ulasan dan memungkinkan model untuk fokus pada kata-kata yang benar-benar menyampaikan sentimen dari ulasan tersebut

Tahap pertama yang dilakukan yaitu proses cleansing. Dalam tahap ini *data set* dilakukan penghapusan atau pembersihan dengan menghilankan tanda baca, angka, karakter, serta stopword. Hasil dapat dilihat pada Tabel 2.

Tabel 2. Sebelum dan Sesudah *Cleansing*

Sebelum	Sesudah
A wonderful little production. The filming technique is very unassuming- very old-time-BBC fashion and gives a comforting, and sometimes discomforting, sense of realism to the entire piece.	A wonderful little production The technique unassuming old time fashion comforting sometimes discomforting sense realism entire piece

Dari hasil tersebut, tag HTML seperti “<br/>” dan stopword seperti “they”, “are”, dan lain-lain sudah berhasil dihapuskan. Hal ini nantinya akan mempermudah model untuk memahami makna dalam setiap kalimat.

Tahap selanjutnya yaitu merupakan proses *normalization.* Dalam langkah ini, setiap kata diubah menjadi bentuk huruf kecil untuk memastikan bahwa model tidak membedakan kata yang sama hanya karena perbedaan kapitalisasi. Lemmatization juga menjadi bagian dari tahap ini, dimana kata-kata tersebut diubah

ke dalam bentuk dasar atau *lemma*-nya. Seperti contoh, kata "comforting" diubah menjadi "comfort", sehingga keduanya dianggap sebagai kata yang sama oleh sistem. Hasil dari proses ini, dapat dilihat pada Tabel 3. Pengurangan variasi kata akan membantu mengurangi jumlah fitur yang perlu diproses oleh model, yang secara langsung akan meningkatkan efisiensi komputasi.

Tabel 3. Sebelum dan Sesudah *Normalization*

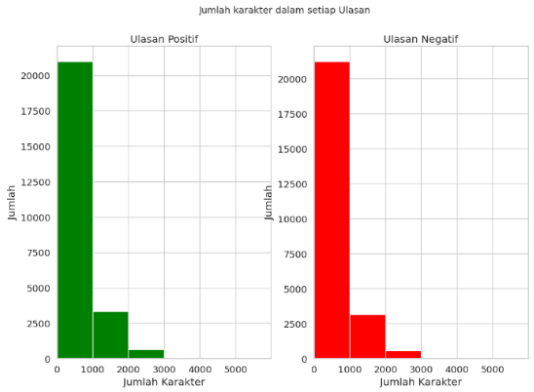
Sebelum	Sesudah
A wonderful little production The technique unassuming old time fashion comforting sometimes discomforting sense realism entire piece	wonderful little production technique unassuming old time fashion comfort sometimes discomforting sense realism entire piece

Terakhir, terdapat tahap *tokenization*, dimana kalimat dirubah menjadi dalam bentuk *token*, seperti yang digambarkan dalam Tabel 4.

Tabel 4. Sebelum dan Sesudah *Tokenization*

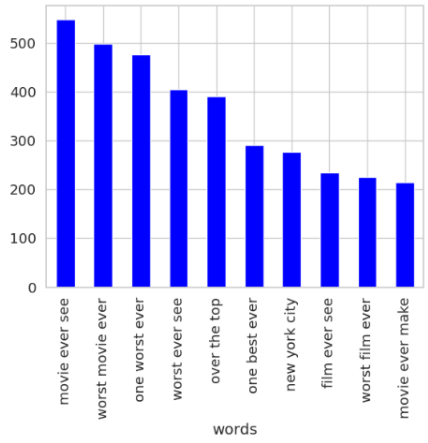
Sebelum	Sesudah
wonderful little production technique unassuming old time fashion comfort sometimes discomforting sense realism entire piece	‘wonderful’, ‘little’, ‘production’, ‘technique’, ‘unassuming’, ‘old’, ‘time’, ‘fashion’, ‘comfort’, ‘sometimes’, ‘discomforting’, ‘sense’, ‘realism’, ‘entire’, ‘piece’

Setelah menyelesaikan tahapan sebelumnya, dilakukan *exploration* kembali untuk melihat perbandingan sebelum dan sesudah *preprocessing text.* Tujuannya untuk mengidentifikasi perubahan yang terjadi akibat dari tahap *preprocessing text.*



Gambar 5. Jumlah Karakter Tiap Sentimen Sesudah *Preprocessing text*

Dengan melihat kepada Gambar 5, diketahui bawa telah terjadi penurunan jumlah karakter. Hal ini menunjukkan bahwa proses sebelumnya telah berhasil. Akibatnya, data yang didapat saat ini lebih bersih dan mencerminkan esensi dari setiap ulasan dengan lebih akurat, serta memungkinkan model untuk menginterpretasi sentimen dengan lebih efektif



Gambar 6. Sepuluh Trigram Dengan Frekuensi Terbanyak Setelah *Preprocessing text*

Kemudian pada Gambar 6, komposisi sepuluh besar trigram mengalami perubahan. Perubahan tersebut antara lain tidak ada tag HTML lagi dalam *data set* dan trigram yang muncul memiliki makna dan dapat dipahami

3.3 BiLSTM

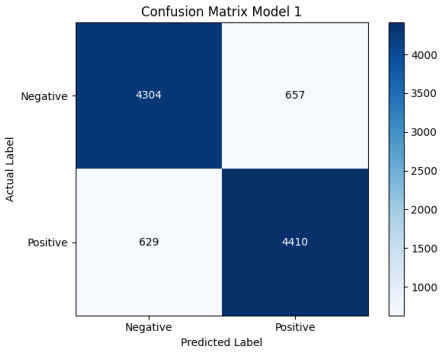
Setelah teks melalui *preprocessing*, kemudian data tersebut dilatih menggunakan model BiLSTM. Dalam penelitian ini, pelatihan model akan dilakukan dalam beberapa skenario. Parameter tetap (Test Size = 0,2; BiLSTM unit = 64; Dropout_1 = 0,3; Dense_1 = 64; Dropout_2 = 0,3; Dense_2 = 2 dengan activation = sigmoid) dan skenario dapat dilihat pada tabel 5.

Tabel 5. Skenario Pengujian

Model	Embedding Dimension	Batch Size
1	32	32
2	32	64
3	64	32
4	64	64

3.4 Evaluasi

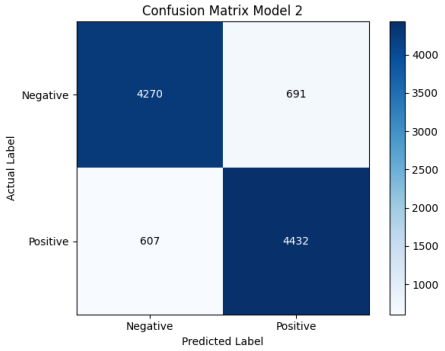
Setelah melalui proses pemodelan dengan beberapa skenario berbeda untuk keempat model, maka dilakukan evaluasi model dengan menggunakan *confusion matrix*.



Gambar 7. *Confusion Matrix* Model 1

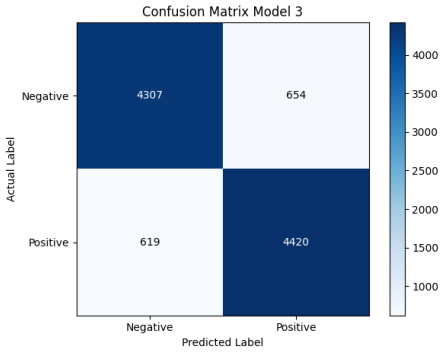
Pada Gambar 7, dapat dilihat bahwa model 1 memiliki nilai *True Negatif* (TN) sebesar 4304, *False Positif* (FP)

sebesar 657, *False Negatif* (FN) sebesar 629, dan *True Positif* (TP) sebesar 4410.



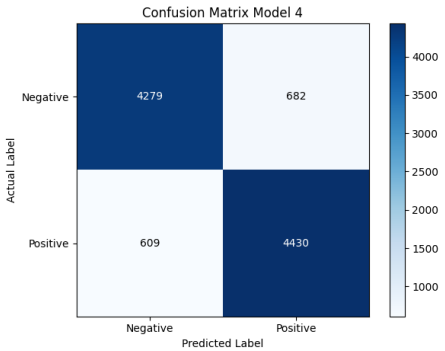
Gambar 8. *Confusion Matrix* Model 2

Confusion matrix model 2 yang dapat dilihat pada Gambar 8 menunjukkan bahwa model 2 memiliki nilai *True Negatif* (TN) sebesar 4270, *False Positif* (FP) sebesar 691, *False Negatif* (FN) sebesar 607, dan *True Positif* (TP) sebesar 4432.



Gambar 9. *Confusion Matrix* Model 3

Pada Gambar 9, model 3 memiliki *confusion matrix* dengan nilai *True Negatif* (TN) sebesar 4307, *False Positif (*FP) sebesar 654, *False Negatif* (FN) sebesar 619, dan *True Positi*f (TP) sebesar 4420.



Gambar 10. *Confusion Matrix* Model 4

Sementara itu, Gambar 10 menunjukkan bahwa model 4 memiliki nilai *True Negati*f (TN) sebesar 4279, *False Positif* (FP) sebesar 682, *False Negatif* (FN) sebesar 609, dan *True Positif* (TP) sebesar 4430.

Berdasarkan *confusion matrix* dari ke empat model, maka dapat dihitung *accuracy*, *precision*, *recall*, dan *F1-Score* yang dapat dilihat pada Tabel 6.

Tabel 6. Hasil Evaluasi Model

Model	Accuracy	Precision	Recall	F1-score
1	0,8714	0,8703	0,8752	0,8727
2	0,8702	0,8651	0,8795	0,8723
3	0,8727	0,8711	0,8772	0,8741
4	0,8709	0,8666	0,8792	0,8728

Berdasarkan hasil evaluasi model, maka didapatkan bahwa Model 3 memiliki Accuracy tertinggi (0,8727) dan juga memiliki F1-Score tertinggi (0,8741) diantara keempat model lainnya. Kombinasi dari Accuracy tertinggi dan F1-Score tertinggi menunjukkan bahwa Model 3 adalah model terbaik di antara keempatnya. Dengan kata lain, Model 3 cenderung memiliki keseimbangan yang baik antara Precision dan Recall (sebagaimana diukur dari F1-Score) serta memprediksi keseluruhan klasifikasi dengan lebih akurat dibandingkan dengan model lainnya.

4. Kesimpulan

Dalam penelitian ini, penelitian dilakukan pada data yang telah melalui tahapan *preprocessing text* dan selanjutnya dilatih menggunakan model BiLSTM. Pelatihan model dijalankan berdasarkan empat skenario dan dilakukan evaluasi menggunakan *Confusion Matrix* dan didapatkan bahwa Model 3 menonjol sebagai model dengan kinerja terbaik, mencapai *accuracy* 0,8727 dan *F1-score* 0,8741. Model 2 menjadi model dengan kinerja terburuk dengan *accuracy* 0,8702 dan *F1-score* 0,8723. Dengan demikian, berdasarkan hasil penelitian ini, Model 3 dengan *Embedding Dimension* 64 dan *Batch Size* 32 memiliki performa terbaik dibandingkan dengan Model lain yang diujikan pada penelitian ini, walaupun secara keseluruhan accuracy tiap model tidak memiliki perbedaan secara signifikan.

Diharapkan pada penelitian berikutnya dapat dikembangkan dengan menambahkan aspek yang dibahas dalam ulasan, menambahkan jumlah *data set* agar kata yang dilatih menjadi lebih beragam, dan menggunakan beberapa model lain sebagai pembanding.

Daftar Rujukan

[1] Md. R. Haque, S. Akter Lima, dan S. Z. Mishu, “Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews,” dalam *2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, IEEE, Des 2019, hlm. 161–164. doi: 10.1109/ICECTE48615.2019.9303573.

[2] K. Sailunaz, M. Dhaliwal, J. Rokne, dan R. Alhajj, “Emotion detection from text and speech: a survey,” *Soc Netw Anal Min*, vol. 8,

no. 1, hlm. 28, Des 2018, doi: 10.1007/s13278-018-0505-2.

[3] B. Pang, L. Lee, dan S. Vaithyanathan, "Thumbs up?," dalam *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP '02*, Morristown, NJ, USA: Association for Computational Linguistics, 2002, hlm. 79–86. doi: 10.3115/1118693.1118704.

[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, dan J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," dalam *Advances in Neural Information Processing Systems*, C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, dan K. Q. Weinberger, Ed., Curran Associates, Inc., 2013. [Daring]. Tersedia pada: https://proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

[5] A. Saxena dan T. Sukumar, "Predicting bitcoin price using lstm And Compare its predictability with arima model," 2018. [Daring]. Tersedia pada: https://api.semanticscholar.org/CorpusID:220374761

[6] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, dan J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, hlm. 68–75, Mar 2017, doi: 10.1049/iet-its.2016.0208.

[7] S. Hochreiter dan J. Schmidhuber, "Long Short-Term Memory," *Neural Comput*, vol. 9, no. 8, hlm. 1735–1780, Nov 1997, doi: 10.1162/neco.1997.9.8.1735.

[8] F. A. Gers, J. Schmidhuber, dan F. Cummins, "Learning to forget: continual prediction with LSTM," dalam *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, 1999, hlm. 850–855 vol.2. doi: 10.1049/cp:19991218.

[9] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad Khasmakhi, M. Asgari-Chenaghlu, dan J. Gao, "Deep Learning Based Text Classification: A Comprehensive Review." November 2020.

[10] M. Rhanoui, M. Mikram, S. Yousfi, dan S. Barzali, "A CNN-BiLSTM Model for Document-Level Sentiment Analysis," *Mach Learn Knowl Extr*, vol. 1, no. 3, hlm. 832–847, Sep 2019, doi: 10.3390/make1030048.

[11] A. Graves dan J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5–6, hlm. 602–610, Jul 2005, doi: 10.1016/j.neunet.2005.06.042.

[12] A. Rolangon *dkk.*, "Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19 The Comparison of LSTM Algorithms for Twitter User Sentiment Analysis on Hospital Services During the Covid-19 Pandemic."

[13] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, dan C. Potts, "Learning Word Vectors for Sentiment Analysis," dalam *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, dalam HLT '11. USA: Association for Computational Linguistics, 2011, hlm. 142–150.

[14] M. T. Anjasmoros, Istiadi, dan F. Marisa, "Seminar Nasional Hasil Riset Prefix-RTR ANALISIS SENTIMEN APLIKASI GO-JEK MENGGUNAKAN METODE SVM DAN NBC (STUDI KASUS: KOMENTAR PADA PLAY STORE)," dalam *Conference on Innovation and Application of Science and Technology (CIASTECH 2020)*, 2020.

[15] I. G. C. P. Yasa, N. A. Sanjaya ER, dan L. A. A. Rahning Putri, "Sentiment Analysis of Snack Review Using the Naïve Bayes Method," *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 8, no. 3, hlm. 333, Jan 2020, doi: 10.24843/JLK.2020.v08.i03.p16.

[16] Z. Huang, W. Xu, dan K. Yu, "Bidirectional LSTM-CRF Models for Sequence Tagging," Agu 2015.

[17] A. Tripathy, A. Agrawal, dan S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Comput Sci*, vol. 57, hlm. 821–829, 2015, doi: https://doi.org/10.1016/j.procs.2015.07.523.