

Analisis Pengaruh *Influencer* Terhadap Produk Yang Dijual Menggunakan Metode *Long Short Term Memory*

Hendrian Sofu Maruhawa, Muhamad Hadi Arfian*

Teknik Informatika, Universitas Esa Unggul, Jakarta
*muhamad.arfian@esaunggul.ac.id

Abstrak— Media sosial merupakan salah satu platform yang banyak digunakan oleh masyarakat untuk berinteraksi, berbagi informasi, dan mengikuti perkembangan terkini. Media sosial juga menjadi sarana pemasaran yang efektif bagi para pelaku bisnis, khususnya dalam bidang produk konsumen. Salah satu strategi pemasaran yang umum di media sosial adalah memanfaatkan *influencer*. Penelitian ini mengeksplorasi dan menganalisis pengaruh *influencer* terhadap produk yang dijual dengan menggunakan teknik crawling data dan metode Long Short Term Memory (LSTM). Data dikumpulkan dari platform Twitter untuk mengidentifikasi pola perilaku konsumen terkait pemasaran produk melalui *influencer*, berdasarkan komentar positif dan negatif. Teknik crawling data digunakan untuk mengumpulkan data yang relevan dari platform online, sedangkan metode LSTM digunakan untuk memodelkan pola temporal yang kompleks. Hasil penelitian menunjukkan bahwa dari perbandingan tiga model LSTM, model LSTM 1 lapisan mencapai akurasi sebesar 63%, model LSTM 2 lapisan mencapai akurasi sebesar 74%, dan model BiLSTM mencapai akurasi sebesar 76%. Dengan dataset yang telah dibersihkan sebanyak 4,353 data, terdiri dari 1,595 data berlabel "Positif" dan 2,758 data berlabel "Negatif", serta menggunakan metode Bi-LSTM dengan perbandingan 30% data testing dan 70% data training, diperoleh akurasi sebesar 77% pada data training. Selain itu, nilai recall sebesar 63%, presisi sebesar 70%, dan nilai f1-score sebesar 66% juga dicapai.

Kata Kunci—Analisis Sentimen; Crawling Data; *Influencer*; LSTM; Produk

DOI: 10.22441/jitkom.v10i1.001

Article History:

Received: Nov 24, 2024

Revised: July 2, 2025

Accepted: Oct 15, 2025

Published: Jan 31, 2026

I. PENDAHULUAN

Media sosial merupakan salah satu platform yang banyak digunakan oleh masyarakat untuk berinteraksi, berbagi informasi, dan mengikuti perkembangan terkini. Media sosial juga menjadi sarana pemasaran yang efektif bagi para pelaku bisnis, khususnya dalam bidang produk konsumen. Salah satu strategi pemasaran yang umum di media sosial adalah memanfaatkan *influencer*. Dalam pemasaran, *influencer* adalah strategi pemasaran yang melibatkan pembuatan dan distribusi konten komersial dengan informasi lisan atau tulisan kepada pelanggan melalui platform seperti blog, website, sosial media, dan video menggunakan individu yang memiliki pengaruh atau kredibilitas di media sosial untuk mempromosikan barang atau jasa kepada audiens mereka[1]. Dalam situasi seperti ini, penting untuk memahami sejauh mana pengaruh *influencer* dapat memengaruhi produk yang mereka promosikan. Diperlukan Analisis sentimen untuk mengidentifikasi dan mengklasifikasikan opini atau polaritas dalam data teks. Saat ini, orang banyak berbagi pendapat dan perasaan mereka di situs sosial, sehingga banyak data dihasilkan secara online[2]. Namun, meskipun popularitas metode ini meningkat, masih terdapat sejumlah permasalahan yang perlu ditangani untuk memahami secara menyeluruh pengaruh *influencer* terhadap produk yang dijual. Permasalahan utama yang mendasari penelitian ini meliputi: Kurangnya Pemahaman tentang pengaruh *influencer*, keterbatasan data dan kurangnya analisis

tren produk. *Machine Learning* yang dipakai adalah Metode deep learning menggunakan representasi data yang lebih tinggi dan abstrak[3]. Peneliti memilih *machine learning* (ML) dikarenakan paling populer untuk menangani set data[4].

II. LITERATURE REVIEW

Adapun penelitian sebelumnya yang pernah dilakukan terkait klasifikasi sentiment dengan judul "Sentiment analysis of novel review using long short-term memory method" dengan data review novel berbahasa Indonesia yang diambil dari situs goodreads.com menunjukkan bahwa metode Long Short-Term Memory (LSTM) memberikan hasil akurasi yang lebih baik dibandingkan dengan metode Naïve Bayes. Metode LSTM mencapai akurasi 72,85%, presisi 73%, recall 72%, dan f-measure 72%. [5]. Dalam studi berjudul "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia" penelitian ini membandingkan algoritma *Bidirectional Long Short-Term Memory* (BiLSTM) dengan algoritma lain (LSTM, *Multinomial Naïve Bayes*, *Logistic Regression*, dan *Support Vector Classifier*) untuk analisis sentimen terhadap pelayanan Grab Indonesia. Penelitian ini menemukan bahwa BiLSTM menghasilkan akurasi terbaik (91%) [6]. Studi serupa dengan judul "Sentiment Analysis of Social Media Users Using Long-Short Term Memory Method" Hasil artikel pengujian metode LSTM menunjukkan kemampuan untuk secara otomatis mengklasifikasikan sentimen tweet pengguna Twitter dengan tingkat akurasi 77% [7]. Peneliti memakai TensorFlow untuk mengembangkan dan menerapkan pembelajaran mesin serta algoritma lain yang memerlukan banyak operasi matematika. [8]. Dengan tools Colab yang

memungkinkan pengguna membuat, menjalankan, dan membagikan notebook Jupyter, yaitu dokumen interaktif yang berisi kode, teks naratif, visualisasi, dan hasil eksekusi kode [9].

III. METODOLOGI PENELITIAN

Analisis sentiment memiliki tahapan perencanaan penelitian. Proses penulisan digambarkan pada Gambar 1 berikut:



Gambar 1. Metode Penulisan

A. Crawling

Tahapan pertama mengumpulkan data. Metode *crawling* ini dapat mengumpulkan data dari aplikasi Twitter dalam bentuk teks tanpa perlu melakukan wawancara atau survei langsung di lapangan [10].

B. Pre-processing

Pada tahap ini Data ulasan yang terkumpul menjalani proses pre-processing Untuk membuat data mentah dapat digunakan dalam pelatihan model pembelajaran mesin, data tersebut perlu diubah menjadi bentuk yang dapat dipahami dan bermanfaat [11]. Berikut tahapan dalam *pre-processing* :

- *Stopword Removal* : Menghapus kata-kata berdasarkan daftar stoplist.
- *Case Folding* : Mengubah teks dalam dokumen menjadi bentuk standar, yaitu lowercase.
- *Tokenisasi (Tokenization)*: Memecah teks menjadi unit kata atau token.
- *Stemming word* : Menguraikan kata menjadi bentuk dasar.

C. Labelling

Labeling dilakukan setelah tahap pre-processing data. Metode labeling ini menggunakan pendekatan berbasis kamus (lexicon-based) yang berisi bobot dari masing-masing kata. Bobot ini digunakan sebagai referensi untuk menghitung nilai sentimen dari kata-kata dalam perhitungan [12].

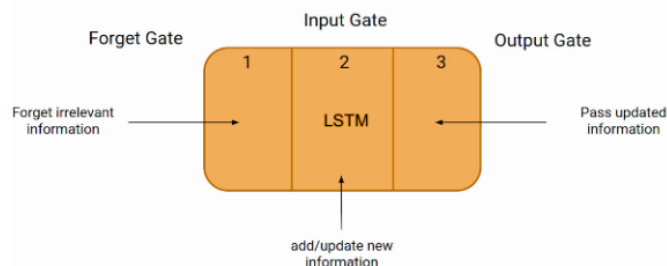
D. Word Embedding

Word embedding merupakan representasi kata-kata dalam bentuk vektor bernilai riil yang mengandung makna semantik dan sintaksis [13]. Dalam penelitian ini, digunakan embedding pra-terlatih seperti Word2Vec.

E. LSTM Model Dan BI-LSTM

Dalam penelitian ini, LSTM model akan dilatih dengan 2 lapisan dengan membandingkan Dengan BI-LSTM

LSTM, menggunakan mekanisme yang efisien untuk memprediksi variabel target dan menentukan bahwa elemen keadaan yang dikodekan akan ditransmisikan ke unit berikutnya setiap saat. Ini mencakup input, output, dan forget gate seperti gambar Ini memungkinkan jaringan untuk belajar urutan lebih lama, mengelola ketergantungan yang lebih lama, dan menemukan solusi tertentu [14].



Gambar 2. LSTM Model

Gerbang yang ada pada satu sel memori Long Short Term Memory (LSTM) dijelaskan sebagai berikut:

Forget Gate adalah gerbang pengganti. Untuk memutuskan apakah input dan output akan diteruskan ke status sel dengan rumus :

$$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f) \quad (1)$$

Gerbang Input memiliki dua fungsi aktivasi, sigmoid dan tanh, untuk memilih bagian yang akan diperbarui, dengan rumus :

$$i_t = \sigma(W_i [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{c}_t = \tanh(W_c [h_{t-1}, x_t] + b_c) \quad (3)$$

Update Cell State gerbang status sel untuk mengkombinasikan forget gate dan input gate untuk menghasilkan gerbang sel yang diperbarui, dengan rumus :

$$C_t = (f_t * C_{t-1} + i_t * \tilde{c}_t) \quad (4)$$

Output Gate yang menentukan bagian mana dari status sel yang akan digunakan untuk menghasilkan hidden state (ht). Output gate berfungsi untuk mengontrol informasi apa yang akan dikeluarkan pada setiap langkah waktu. Nilai baru dan lama digabungkan dalam gerbang keluaran pertama, dengan menggunakan dua rumus :

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t * \tanh(C_t) \quad (5)$$

Karena LSTM dengan 2 lapisan memiliki kelemahan yaitu hanya memproses kata dalam satu arah, dari awal hingga akhir, penulis mempertimbangkan penggunaan Bidirectional LSTM. Dua lapisan tersebut bekerja dalam arah yang berlawanan dan dapat memprediksi informasi baik dari masa lalu maupun masa depan dengan menerapkan forward pass dan backward pass [15]. Bidirectional LSTM memiliki dua lapisan LSTM yang

terpisah: satu untuk arah maju dan satu untuk arah mundur, dengan rumus:

$$h_t^{BiLSTM} = h_t^{forward} \oplus h_t^{backward} \quad (6)$$

F. Evaluasi

Dalam tahap evaluasi penulis menghitung performa klasifikasi menggunakan Confusion Matrix. Confusion Matrix biasanya berupa tabel yang digunakan untuk mengukur kinerja suatu proses atau model klasifikasi, dengan mencatat nilai true positif, false positif, true negatif, dan false negatif. Berikut merupakan tabel Confusion matrix :

Tabel 1. Confusion Matrix

Kelas Sebenarnya	Kelas Prediksi	
	Negatif	Positif
Negatif	True Negatif (TN)	False Positif (FP)
Positif	False Negatif (FN)	True Positif (TP)

Persamaan ini memungkinkan kita untuk menghitung berbagai metrik evaluasi yang penting untuk mengukur performa model klasifikasi sebagai berikut :

$$Akurasi (\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (7)$$

$$Precision (\%) = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$Recall (\%) = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

$$F1 - Score (\%) = 2 \times \frac{precision \times recall}{precision + recall} \times 100\% \quad (10)$$

IV. HASIL DAN ANALISA

A. Crawling

Data yang dikumpulkan mencakup tweet yang menyebutkan endorse atau influencer sebanyak 10 influencer dalam periode waktu 2018 - 2024. semua hasil crawling menggunakan berbagai query yang relevan dan dari tahun yang berbeda, penulis menggabungkan semua file tersebut menjadi satu dan memperoleh 4.745 data. dan kolom full_text akan digunakan untuk proses analisis.

Gambar 3. Hasil Crawling

B. Pre-processing data

Dari data crawling yang sebelumnya didapat 4,745 menjadi 4,354 setelah dilakukannya preprocessing. Gambar 4 menunjukkan hasil sebelum dilakukan pre-processing :

Gambar 4. Sebelum Pre-processing

Proses ini meliputi beberapa tahapan penting seperti penghapusan stopwords, case folding, tokenisasi dan stemming. Tujuan dari pre-processing ini adalah untuk membersihkan dan mempersiapkan data agar lebih relavan dan siap digunakan dalam analisis lebih lanjut. Berikut hasil setelah pre-processing:

Gambar 5. Setelah Pre-processing

C. Labeling

Data yang telah melalui proses pre-processing akan diberi label secara otomatis untuk mengidentifikasi kelas sentimen menggunakan kamus (lexicon). Dalam analisis sentimen ini, tujuannya adalah menentukan kelas sentimen positif dan negatif. pelabelan data untuk analisis yang lebih mendalam. Berikut hasil data Positif diberi label karena mengandung kata-kata yang menggambarkan kepuasan, penerimaan, atau pengalaman positif terhadap hal yang diulas. Sebaliknya, Data Negatif cenderung diberi label tersebut karena terdapat kata-kata yang mencerminkan kekecewaan, ketidakpuasan, atau kritik terhadap sesuatu yang direview.

```

5 Data Positif:
year          stemmed_text sentiment
0 2018 ['awkarin', 'pp', 'review', 'hari', 'dapet', '...' positive
1 2018 ['lg', 'dua', 'dian', 'trs', 'drted', 'gue', 'm...' positive
3 2018 ['review', 'salah', 'produk', 'beli', 'brand', '...' positive
4 2018 ['suka', 'video', 'review', 'bad', 'influence'... positive
5 2018 ['kakkakak', 'semuaaa', 'yg', 'kasi', 'saran... positive

5 Data Negatif:
year          stemmed_text sentiment
2 2018 ['video', 'playlist', 'review', 'bad', 'influe... negative
6 2018 ['review', 'ternak', 'meme', 'suka', 'bangga'... negative
8 2018 ['awkarin', 'awkarin', 'beda', 'sih', 'review'] negative
9 2018 ['yg', 'awkarin', 'review', 'kalo', 'gasalah',... negative
10 2018 ['akun', 'review', 'hp', 'gue', 'tau', 'awarki... negative
    
```

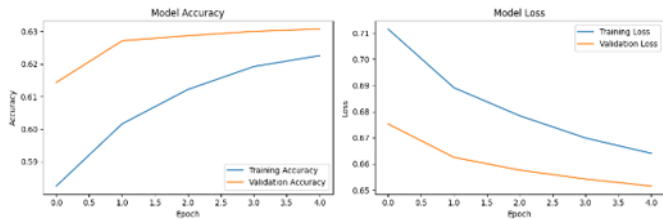
Gambar 6. Labeling

D. Word Embedding

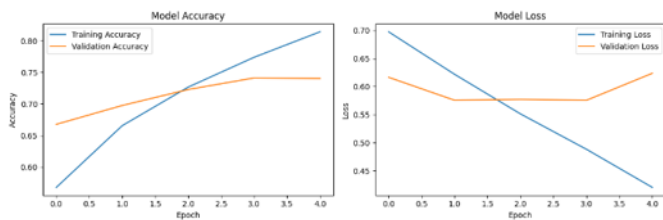
Setelah selesai melakukan labeling penulis melanjutkan tahapan word embedding menggunakan metode word2vec Dengan 300 dimensi untuk menghasilkan representasi kata dalam bentuk vektor dense yang lebih efisien.

E. Pengujian LSTM dan BI-LSTM

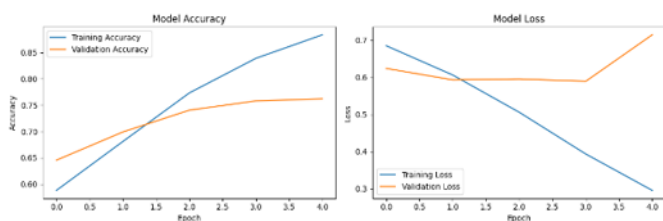
Berdasarkan penelitian sebelumnya peneliti menguji data menggunakan 100 neuron untuk memastikan kapasitas yang cukup dalam menangkap pola kompleks tanpa risiko overfitting. Dropout sebesar 0,9 diterapkan untuk mencegah overfitting dengan menonaktifkan 90% neuron selama pelatihan. Optimizer Adam dengan learning rate 0,001 digunakan untuk menjamin konvergensi model yang stabil, sementara 5 epoch dan aktivasi sigmoid dipilih untuk menghasilkan output antara 0 dan 1. Hasil pengujian ini kemudian dibandingkan antara model LSTM dengan 1 lapisan, LSTM dengan 2 lapisan, dan BI-LSTM dalam menghasilkan kinerja model yang ditampilkan dalam bentuk grafik sebagai berikut :



Gambar 7. Grafik LSTM 1 Lapisan



Gambar 8. Grafik LSTM 2 Lapisan



Gambar 9. Grafik BI-LSTM

Setelah melakukan pengujian LSTM dengan 1 lapisan dan 2 lapisan menggunakan BI-LSTM, hasil menunjukkan bahwa BI-LSTM memiliki nilai akurasi yang tinggi. Oleh karena itu, selanjutnya dibuat model arsitektur untuk memanfaatkan representasi vektor kata yang kaya dari embedding layer, dengan hasil arsitektur model BI-LSTM.

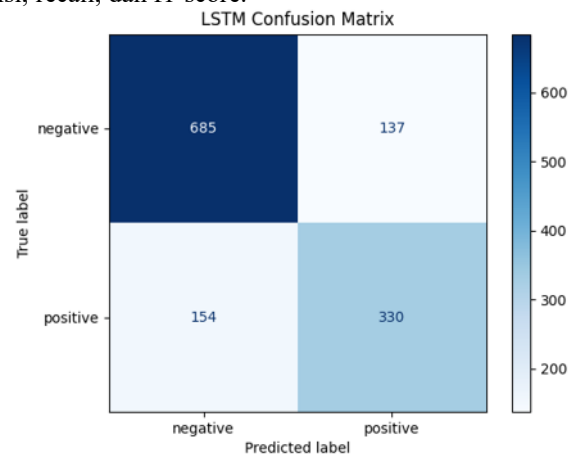
Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 47, 300)	3145200
bidirectional (Bidirectional)	(None, 47, 200)	328800
dropout (Dropout)	(None, 47, 200)	0
bidirectional_1 (Bidirectional)	(None, 200)	248800
dropout_1 (Dropout)	(None, 200)	0
dense (Dense)	(None, 1)	201

Total params: 3707001 (14.14 MB)
 Trainable params: 561801 (2.14 MB)
 Non-trainable params: 3145200 (12.00 MB)

Gambar 11. Model Arsitektur

F. Evaluasi

Pada tahap ini, confusion matrix berukuran 2 x 2 akan digunakan untuk mengevaluasi kinerja algoritma LSTM dengan menampilkan jumlah prediksi yang benar dan salah yang dihasilkan oleh model dibandingkan dengan nilai aktual yang diketahui. Ini akan memungkinkan perhitungan akurasi, presisi, recall, dan f1-score.



Gambar 12. Hasil Confusion Matrix

Dari confusion matrix yang ditampilkan dapat digunakan untuk menghitung akurasi, presisi, recall, dan f1-score sebagai bagian dari proses validasi nilai evaluasi.

$$Akurasi (\%) = \frac{1014}{1306} \times 100\% = 77\%$$

$$Precision (\%) = \frac{330}{467} \times 100\% = 70\%$$

$$Recall (\%) = \frac{309}{484} \times 100\% = 63\%$$

$$F1 - Score (\%) = 2 \times 0.33 \times 100 = 66\%$$

V. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, dapat ditarik kesimpulan bahwa nilai akurasi sebesar 77 % setelah mendapatkan hasil akurasi didapatkan nilai Recall yang dihasilkan 63 %, Presisi sebesar 70 % dan nilai f1-score 66 %. Metode Long short-term memory (LSTM) layak digunakan untuk klasifikasi sentimen pengaruh influencer terhadap penjualan produk dengan hasil sentimen negatif berupa

ketidakpuasan terhadap influencer dalam merekomendasi penjualan produk adapun saran penelitian menggunakan metode lain seperti KNN, Naive bayes dan menggunakan selain bahasa Indonesia.

DAFTAR PUSTAKA

- [1] Ş. Yazgan Pektaş and A. Tunç Hassan, "The Effect of Digital Content Marketing on Tourists' Purchase Intention," *Journal of Tourismology*, vol. 6, no. 1, pp. 79–98, June 2020, doi: 10.26650/jot.2020.6.1.0011.
- [2] N. Parveen, P. Chakrabarti, B. T. Hung, and A. Shaik, "Twitter sentiment analysis using hybrid gated attention recurrent network," *Journal of Big Data*, vol. 10, no. 1, Apr. 2023, doi: 10.1186/s40537-023-00726-3.
- [3] J. Hwang, J. Seo, J. Kim, S. Park, Y. Kim, and K. Kim, "Comparison between Deep Learning and Conventional Machine Learning in Classifying Iliofemoral Deep Venous Thrombosis upon CT Venography," *Diagnostics*, vol. 12, no. 2, p. 274, Jan. 2022, doi: 10.3390/diagnostics12020274.
- [4] E. G. Ceja, "Behavior Analysis with Machine Learning Using R," Oct. 2021, doi: 10.1201/9781003203469.
- [5] M. A. Nurrohmat and A. SN, "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 3, p. 209, July 2019, doi: 10.22146/ijccs.41236.
- [6] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 12, no. 2, pp. 89–99, Sept. 2022, doi: 10.34010/jamika.v12i2.7764.
- [7] C. K. N. Paputungan and A. Jacobus, "Analisis Sentimen Pengguna Sosisal Media Menggunakan Metode Long Short Term Momory," *Jurnal Teknik Elektro dan Komputer*, vol. 10, no. 2, pp. 99–106, 2021, doi: 10.35793/jtek.v10i2.32893.
- [8] Jarvin, "Apa Itu Tensorflow?," Binus. Accessed: Aug. 01, 2024. [Online]. Available: <https://sis.binus.ac.id/2024/01/18/apa-itu-tensorflow/>
- [9] K. K. Handaya and S. Wahyu, "Sentiment Analysis on Hate Speech Post 2024 Election for Elected President Using a Hybrid Model Machine Learning," *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, vol. 10, no. 2, Dec. 2024, doi: 10.24014/coreit.v10i2.31927.
- [10] R. Tineges, "Teknik Pengumpulan Data Sekunder dengan Web Crawling," dqlab. Accessed: Aug. 01, 2024. [Online]. Available: <https://dqlab.id/teknik-pengumpulan-data-sekunder-dengan-web-crawling>
- [11] H. A. Madni *et al.*, "Improving Sentiment Prediction of Textual Tweets Using Feature Fusion and Deep Machine Ensemble Model," *Electronics*, vol. 12, no. 6, p. 1302, Mar. 2023, doi: 10.3390/electronics12061302.
- [12] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5543–5574, Jan. 2022, doi: 10.1007/s10462-021-10134-9.
- [13] B. Wang, A. Wang, F. Chen, Y. Wang, and C. C. J. Kuo, "Evaluating word embedding models: Methods and experimental results," *APSIPA Trans. Signal Inf. Process.*, vol. 8, pp. 1–14, 2019, doi: 10.1017/ATSIP.2019.12.
- [14] T. Mamo and F. K. Wang, "Long short-term memory with attention mechanism for state of charge estimation of lithium-ion batteries," *IEEE Access*, vol. 8, pp. 94140–94151, 2020, doi: 10.1109/ACCESS.2020.2995656.
- [15] S. Ahmed *et al.*, "Att-bil-sl: Attention-based bi-lstm and sequential lstm for describing video in the textual formation," *Appl. Sci.*, vol. 12, no. 1, pp. 1–19, 2022, doi: 10.3390/appl2010317.