

Neural Machine Translation Sunda Halus - Loma Menggunakan Algoritma LSTM Dengan Optimizer ADAM

Marsela Arsyah Sakinah, Teguh Ikhlas Ramadhan, Rudi Hartono
Prodi Teknik Informatika / Universitas Perjuangan Tasikmalaya
Tasikmalaya, (0265) 326058/Universitas Perjuangan
e-mail: 2003010086@unper.ac.id

Abstrak

Penelitian ini mengimplementasikan Neural Machine Translation (NMT) dengan algoritma Long Short Term Memory (LSTM) untuk meningkatkan kualitas terjemahan Bahasa Sunda Halus ke Bahasa Sunda Loma. Bahasa Sunda merupakan bagian penting dari budaya masyarakat Sunda dengan konsep undak usuk basa yang khas. Meskipun Bahasa Sunda terancam punah, NMT diharapkan dapat membantu melestarikannya. Model NMT berbasis LSTM dilatih dengan optimizer ADAM dan dievaluasi dengan BLEU score. Hasil menunjukkan tingkat kesesuaian yang tinggi antara terjemahan model dengan referensi manusia pada data pelatihan dan data testing. Skor BLEU pada dataset menunjukkan kemampuan model dalam menangkap struktur dan konteks kalimat dengan baik, Bleu skor yang diperoleh sebesar 96,975%. Penelitian ini tidak hanya bertujuan untuk meningkatkan akurasi terjemahan, tetapi juga memperkenalkan kontribusi teknologi NMT dalam pelestarian budaya bahasa daerah. Hasilnya diharapkan dapat menjadi dasar bagi pengembangan teknologi terjemahan otomatis dan pelestarian budaya bahasa daerah di Indonesia.

Kata kunci— NMT, Struktur LSTM, ADAM, Sunda, undak usuk basa

Abstract

This research implements Neural Machine Translation (NMT) with Long Short Term Memory (LSTM) algorithm to enhance the translation quality from Formal Sundanese to Informal Sundanese. Sundanese language is an integral part of Sundanese culture with its unique "undak usuk basa" concept. Despite being endangered, NMT is expected to aid in its preservation. LSTM-based NMT model is trained using ADAM optimizer and evaluated with BLEU score. Results indicate high alignment between model translation and human reference on both training and testing data. BLEU scores demonstrate the model's ability to capture sentence structure and context effectively, achieving a BLEU score of 96.975%. The study aims not only to improve translation accuracy but also to introduce NMT technology's contribution to preserving regional language culture. Findings are expected to serve as a foundation for automatic translation technology development and regional language preservation in Indonesia.

Keywords— NMT, LSTM structure, ADAM, Sundanese, undak usuk basa

1. Pendahuluan

Dalam era globalisasi informasi saat ini, kemampuan untuk mentransfer makna antara berbagai bahasa menjadi semakin penting. Terjemahan mesin, khususnya menggunakan teknologi Neural Machine Translation (NMT), telah menjadi fokus utama dalam pengembangan sistem untuk memfasilitasi komunikasi lintas bahasa dengan lebih efisien dan akurat. Perkembangan *Neural Machine Translation* (NMT) menjadi peran penting dalam bidang pemrosesan bahasa alami dan terjemahan otomatis. Dengan memanfaatkan perkembangan NMT, penelitian ini bertujuan untuk meningkatkan kualitas terjemahan dan meningkatkan kemampuan model dalam menangani Bahasa Sunda. Encoder dan decoder dalam mesin penerjemah memproses dengan pendekatan jaringan syaraf tiruan atau *Recurrent Neural Network* (RNN) yang dikembangkan menjadi NMT. Penggunaan NMT untuk mesin penerjemah menjadi populer dalam beberapa penelitian sebelumnya, Penelitian terkini berfokus pada arsitektur RNN [1].

Dalam penelitian ini, Bahasa Sunda adalah salah satu bahasa daerah yang digunakan oleh masyarakat di wilayah Jawa Barat, Indonesia. Bahasa daerah merupakan bahasa yang digunakan di suatu

wilayah dalam sebuah negara, baik itu daerah kecil, provinsi atau lebih luas [2]. Bahasa Sunda digunakan setidaknya 38 juta orang dan termasuk bahasa ibu dengan penutur terbanyak kedua di Indonesia setelah Bahasa Jawa [3]. Namun seiring dengan perkembangan zaman, masyarakat mulai melupakan bahasa daerahnya. Bahasa Sunda terancam punah, hanya sekitar 40% anak di Jawa Barat yang menggunakan Bahasa Sunda [4]. Bahasa ini memiliki ciri khas tersendiri dan merupakan bagian penting dari identitas budaya masyarakat Sunda yaitu *undak usuk basa*. *Undak usuk basa* adalah suatu sistem penggunaan variasi Bahasa Sunda halus, sedang (loma), dan kasar, sehingga berkesinambungan dengan tata krama kesopanan dalam berkomunikasi untuk saling menghormati [5].

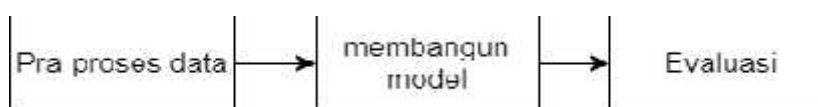
Bahasa Sunda halus terbagi menjadi dua yang digunakan untuk menghormati lawan bicara, yaitu untuk penuturnya sendiri dan untuk orang lain, sedangkan Bahasa Sunda loma tidak membedakan siapa penuturnya karena digunakan untuk lawan bicara yang akrab. Penggunaan Bahasa Sunda halus sering kali sulit dibedakan karena bergantung pada objek, maka kata yang digunakan harus menyesuaikan dengan objek tersebut. Tidak semua masyarakat yang tinggal di daerah Sunda mengetahui perbedaan penggunaan kata bahasa Sunda halus atau Loma, karena ada beberapa kalimat yang jarang digunakan.

NMT memanfaatkan konsep *deep learning* yang kompleks untuk menerjemahkan bahasa. Konsep kerja dalam NMT adalah dengan menerjemahkan suatu bahasa dengan mengumpulkan data dalam bentuk teks, sehingga dapat mengajarkan komputer agar memiliki kemampuan yang dapat menyerupai kemampuan manusia yaitu belajar dari pengalaman. Penelitian ini mengimplementasikan model NMT berbasis algoritma *Long Short Term Memory* (LSTM). LSTM merupakan jenis arsitektur jaringan syaraf tiruan yang digunakan dalam machine learning. LSTM merupakan model jaringan syaraf tiruan berulang yang digunakan untuk memecahkan masalah terhadap ketergantungan urutan panjang pada jaringan syaraf berulang dengan menambahkan unit memori, *input gates*, *output gates*, *forgetting gates*, dan meningkatkan kemampuan jaringan syaraf berulang untuk memproses data [6]. Dengan menggunakan optimizer *Adaptive Moment Estimation* (ADAM) dalam pelatihan model. ADAM dikenal mempunyai kemampuan yang baik dalam mengoptimalkan proses pelatihan jaringan syaraf [7].

Penelitian ini tidak hanya untuk meningkatkan akurasi terjemahan antara bahasa Sunda Halus dan Bahasa Sunda loma, tetapi juga untuk memperkenalkan kontribusi teknologi NMT dalam melestarikan dan mengembangkan warisan budaya bahasa daerah Indonesia. Melalui penelitian ini dan evaluasi yang cermat, kami berharap untuk memperoleh pemahaman yang lebih dalam tentang efektivitas model NMT berbasis LSTM dalam konteks terjemahan bahasa Sunda Halus-Loma. Hasil penelitian ini diharapkan dapat memberikan kontribusi positif bagi pengembangan teknologi terjemahan otomatis dan pelestarian budaya bahasa daerah di Indonesia, serta menjadi dasar untuk penelitian lanjutan dalam bidang ini.

2. Metodologi Penelitian

Penelitian ini digambarkan dalam bentuk diagram alur untuk menggambarkan alur penelitian agar lebih mudah dalam menyampaikan informasi. Tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut.



Gambar 1. Metode penelitian

Penelitian ini melalui persiapan data, pembangunan model, dan evaluasi model terlihat pada Gambar 1. Tahap Pra-proses data merupakan proses mengumpulkan data dan mengubah data teks menjadi token. Membangun model merupakan persiapan bagaimana model dapat mengolah teks dan model dapat melatih sehingga model dapat menyelesaikan masalah penerjemahan. Evaluasi adalah mengevaluasi hasil yang telah dihasilkan oleh model.

2.1 Pra-proses Data

Dataset yang digunakan telah dikumpulkan sebanyak 8000 kalimat yaitu Bahasa Sunda halus dan Sunda loma. Proses pembersihan data terdiri dari data cleaning, yaitu menghilangkan karakter yang tidak diperlukan dan mengubah huruf kapital menjadi huruf kecil menggunakan fungsi `lower()` dalam bahasa pemrograman Python, seperti pada Tabel 1.

Tabel 1. Contoh data set setelah tahap *data cleaning*, dengan bahasa sumber (halus) dan bahasa target (loma)

HALUS	LOMA
abdi hoyong neda	urang hayang dahar
dupi anjeun bade tuang	ari maneh erek dahar
abdi tos neda nembe di rorompok	urang geus dahar bieu di imah
benten deui eta mah di bumi anjeun	beda deui eta mah di imah maneh
enggal tuang heula	geuwat dahar heula

Langkah berikutnya adalah tokenisasi, yang merupakan proses mengonversi data bahasa sumber dan bahasa target menjadi token. Proses tokenisasi ini mengubah kalimat menjadi kata-kata yang akan digunakan untuk membangun kosa kata. Setiap kata dalam teks kemudian akan diberi kode dalam bentuk bilangan bulat (integer).

Tabel 2. Proses tokenisasi (memecah kalimat menjadi kata)

HALUS		LOMA	
Kalimat	Token	Kalimat	Token
abdi hoyong neda		urang hayang dahar	
dupi anjeun bade tuang		ari maneh erek dahar	
abdi tos neda nembe di rorompok		urang geus dahar bieu di imah	
benten deui eta mah di bumi anjeun		beda deui eta mah di imah	
enggal tuang heula		maneh	
		geuwat dahar heula	

Tabel 3. Membuat kode integer pada token

HALUS		LOMA	
Kalimat	Token	Kalimat	Token
abdi	2	urang	22
hoyong	3	hayang	23
neda	4	dahar	24

Tabel 4. Menggabungkan kalimat awal dengan kode token

HALUS		LOMA	
Kalimat	Token	Kalimat	Token
abdi hoyong neda	[2, 3, 4]	urang hayang dahar	[22, 23, 24]
dupi anjeun bade tuang	[5, 1, 6, 7]	ari maneh erek dahar	[25, 21, 26, 27]
abdi tos neda nembe di rorompok	[2,9,4,12,56, 34]	urang geus dahar bieu di imah	[22,34,24,76,45, 56]

Setelah data divektorisasi, data tersebut harus dipecah menjadi set pelatihan dan set pengujian. Training set digunakan untuk melatih model, dan test set digunakan untuk menguji model dan melakukan evaluasi apakah penggunaan model neural ini sudah merupakan model yang baik atau buruk. Data yang digunakan dibagi untuk data pelatihan dan pengujian. Tujuan data pelatihan adalah untuk melatih model saat dalam proses pelatihan kemudian disimulasikan sebagai set pengujian.

2.2 Membangun Model

Membangun model dengan membuat objek dan menambahkan lapisan LSTM. Input dan output didefinisikan, di mana input berupa urutan kata atau karakter dalam bentuk tensor, sementara outputnya adalah urutan kata dalam bahasa target. Selanjutnya, konfigurasi LSTM dilakukan dengan menentukan parameter untuk setiap lapisan model LSTM, termasuk jumlah unit dan fungsi aktivasi. Tahap terakhir adalah pelatihan model, di mana model akan memperbaiki parameter internalnya secara iteratif untuk meminimalkan fungsi kerugian selama proses pelatihan.

Penelitian ini menggunakan *Recurrent Neural Network* yang dimodifikasi sebagai model yaitu *Long Short Term Memory* (LSTM). Model LSTM baik untuk menangkap informasi suatu sequence karena kemampuannya menyimpan sebagian memori sehingga dapat mengatasi masalah vanishing gradien yang terjadi pada Plain Recurrent Neural Network. Input dari model akan berupa kalimat bahasa Inggris yang divektorkan dan outputnya juga akan berupa kalimat bahasa Sunda yang divektorkan. Arsitektur RNN pada model jaringan syaraf tiruan menggunakan arsitektur many-to-many.

2.3 Evaluasi

Skor BLEU adalah matrik evaluasi otomatis dan merupakan singkatan dari *Bilingual Evaluation Understudy* [8]. Bleu digunakan untuk mengetahui sejauh mana terjemahan yang dihasilkan oleh mesin penerjemah mendekati hasil terjemahan manusia. BLEU dihitung dengan menghitung kata-kata dalam hasil terjemahan mesin yang sesuai dengan referensi. Skor BLEU berkisar dari 0 hingga 1 atau 0 hingga 100, dengan 0 menunjukkan tidak ada kecocokan dan 1 menunjukkan semua kecocokan, yang tidak mungkin dilakukan untuk semua kalimat pengujian [9].

Precision lebih cocok digunakan untuk menilai kalimat pendek, sementara dapat menimbulkan pertanyaan tentang keakuratan terjemahan mesin untuk kalimat yang lebih panjang, meskipun tetap mempertahankan tingkat presisi yang tinggi. Untuk mengatasi masalah ini, *brevity penalty* diperlukan agar evaluasi lebih seimbang antara panjang kalimat dalam terjemahan dan referensi. Wn adalah bobot yang diberikan pada presisi n-gram yang telah dimodifikasi.

$$Brevity Penalty (bp) = \begin{cases} 1, & \text{if } c > r, \\ e^{1-(r/c)}, & \text{if } c \leq r, \end{cases}$$

Dimana c merupakan panjang dari calon kalimat dan r adalah panjang kalimat referensi.

$$BLEU = bp \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right)$$

3. Hasil dan Pembahasan

3.1. Pengumpulan data

Total data keseluruhan yang digunakan adalah 8000 untuk Bahasa Sunda halus begitujua untuk Bahasa Sunda loma. Pada Tabel 5. pembagian data sebanyak 90% dari 8000 digunakan untuk melatih model dan 10% -nya digunakan untuk pengujian model.

Tabel 5. Pembagian data testing dan data training.

Total kalimat paralel	Kalimat paralel training	Kalimat paralel testing
8000	7200	800

Panjang kalimat yang ada didalam dataset berbeda beda. Pada Tabel 6. Merupakan kalimat acak dari bahasa sumber (halus) dan bahasa target (loma) dengan jumlah kalimat yang bervariasi. Kemudian pada Tabel 7 merupakan contoh kalimat yang telah dilakukan *data cleaning* yaitu mengubah huruf kapital menjadi kecil dan menghilangkan karakter yang tidak diperlukan.

Tabel 6. Contoh dataset sebelum melalui tahap *data cleaning*

HALUS	LOMA
Nyeuri patuangan saatos abdi neda. tahu ti anjeun patuangan anjeun na alit!	Nyeuri beuteung saenggeus urang dahar. Tahu ti maneh beuteung maneh na lemet!
Miceun heula!! ngiring ka pun bibi	Ngising heula!! milu ka bibi urang

Tabel 7. Contoh dataset setelah melalui proses *data cleaning*

HALUS	LOMA
nyeuri patuangan saatos abdi neda tahu paparin ti anjeun patuangan anjeun na alit miceun heula ngiring ka pun bibi	nyeuri beuteung saenggeus urang dahar tahu pamere maneh beuteung maneh na lemet ngising heula milu ka bibi urang

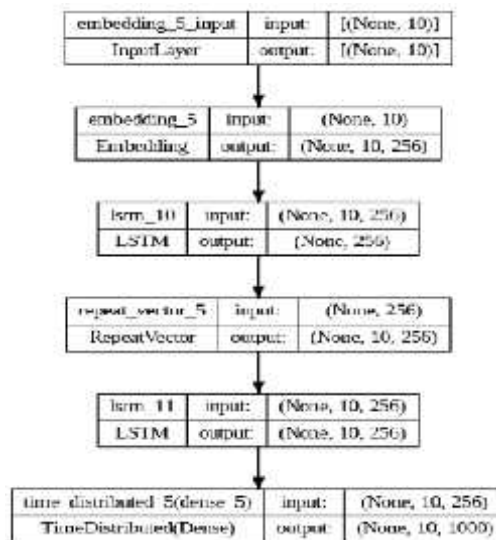
Dihasilkan 1567 encode pada Bahasa sunda Halus dan 1533 untuk Bahasa Sunda loma. Encode merupakan kalimat yang diurutkan dengan kode integer dalam tahap tokenisasi. Berikut adalahh contoh hasil tokenisasi dalam dataset yang digunakan.

Tabel 8. Hasil penggabungan kalimat dengan ecode pada tahap tokenisasi

HALUS		LOMA	
Kalimat	encode	Kalimat	encode
bapak bade angkat damel	[2, 3, 4, 9]	Bapak erek indit gawe	[22, 23, 24, 90]
abdi kirang seneng kana lauk	[12, 56, 8, 21, 65, 17]	Urang kurang beuki kana lauk	[23, 89, 34, 14, 3]
naha anjeun atos mulih ti sawah	[2,9,4,12,56, 34]	urang geus dahar bieu di imah	[22,34,24,76,45, 56]

3. 2 Membangun Model

Lapisan input pertama digunakan untuk membaca vektor input berukuran 10 dari bahasa sumber. Setelah itu, vektor input disematkan ke dalam lapisan embedding sebelum diproses oleh lapisan LSTM. Lapisan LSTM pertama bertindak sebagai encoder, dan outputnya diteruskan ke lapisan RepeatVector. Lapisan RepeatVector berperan sebagai penghubung antara lapisan LSTM pertama dan lapisan LSTM kedua, yang juga bertindak sebagai encoder. Decoder LSTM menerima output dari lapisan LSTM kedua dan menuju ke lapisan dense. Pada lapisan dense, distribusi probabilitas untuk setiap kelas dalam kosakata target dihitung menggunakan aktivasi softmax. Ringkasan model dapat dilihat pada Gambar 2 dan Gambar 3.



Gambar 2. Arsitektur model dimulai dari lapisan input sampai output

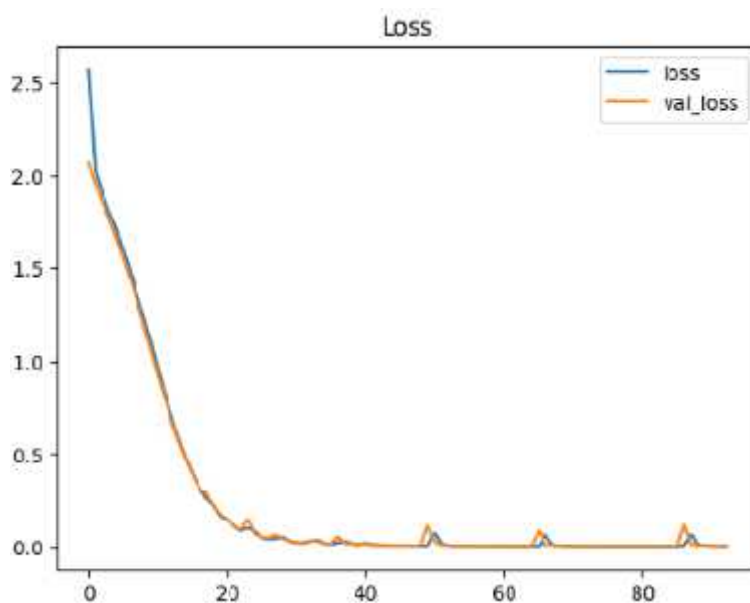
```

Model: "sequential_5"
-----
Layer (type)                Output Shape              Param #
-----
embedding_5 (Embedding)     (None, 10, 256)         256000
lstm_10 (LSTM)              (None, 256)             525312
repeat_vector_5 (RepeatVec  (None, 10, 256)         0
tor)
lstm_11 (LSTM)              (None, 10, 256)         525312
time_distributed_5 (TimeDi  (None, 10, 1000)        257000
stributed)
-----
Total params: 1563624 (5.96 MB)
Trainable params: 1563624 (5.96 MB)
Non-trainable params: 0 (0.00 Byte)
  
```

Gambar 3. Lapisan embedding menerima input dengan ukuran 10, merupakan vektor maksimum dalam bahasa sumber. Lapisan akhir menghasilkan output berukuran 10, merupakan vektor maksimum untuk target.

Arsitektur model Neural Machine Translation (NMT) terdiri dari lapisan Embedding yang mengubah kata-kata dalam urutan input menjadi vektor embedding dengan dimensi 256, yang penting untuk merepresentasikan makna kata-kata secara numerik. Lapisan LSTM dengan 256 unit digunakan untuk memproses vektor embedding dan menghasilkan representasi urutan input dengan dimensi yang lebih rendah. Lapisan RepeatVector mengulang vektor hasil dari LSTM pertama sehingga panjang urutan output menjadi 10 timestep, sesuai dengan struktur bahasa keluaran yang diinginkan. Lapisan LSTM kedua memproses urutan yang telah diulang dan mengembalikan seluruh urutan sebagai output, memungkinkan model memahami dan menangkap pola dalam urutan input yang dapat direplikasi dalam urutan output. Lapisan Dense dengan aktivasi softmax mengonversi representasi numerik urutan output menjadi distribusi probabilitas untuk setiap kata dalam kosakata output, menentukan kemungkinan kata pada setiap posisi dalam urutan output. Model ini memiliki sekitar 1,563,624 parameter total dan dapat disesuaikan selama pelatihan untuk meningkatkan kemampuannya dalam menerjemahkan bahasa.

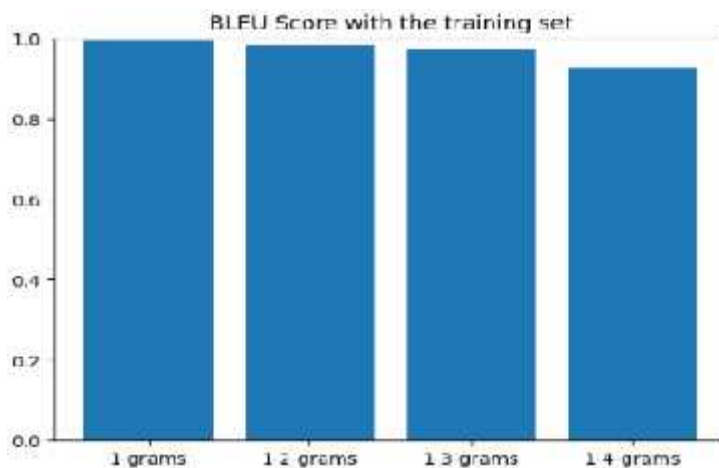
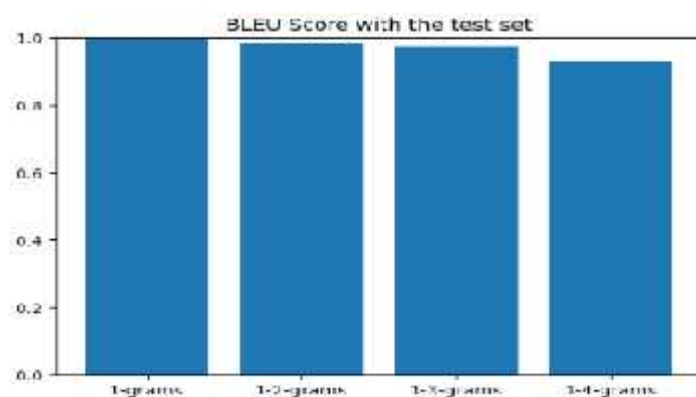
3.3 Evaluasi



Gambar 4. Grafik loss training menggunakan ADAM

Gambar 4. Merupakan hasil dari proses pelatihan menggunakan optimizer ADAM. Grafik menunjukkan bahwa garis train (biru) dan garis validasi (jingga) saling mendekat dan menurun seiring dengan berjalannya epoch. Kedua garis tersebut cenderung stabil dan berhenti berfluktuasi setelah epoch ke-93. Hal ini menandakan bahwa model yang dibangun berhasil dalam proses pelatihan dan generalisasi, karena kedua kurva mendekati konvergensi dengan baik. Penggunaan optimizer ADAM telah membantu menemukan solusi yang baik, karena ADAM memiliki tingkat akurasi yang lebih baik dibandingkan dengan RMSprop dalam banyak kasus [7].

Pelatihan dilakukan menggunakan optimizer Adam dengan loss function categorical crossentropy. Proses pelatihan dilakukan dengan melakukan fit terhadap data *training* untuk Sunda halus dan *training* untuk Sunda loma selama 200 epoch dengan ukuran batch sebesar 64. Sebanyak 10% dari data latih digunakan sebagai data validasi. Selama pelatihan, digunakan callback EarlyStopping untuk memantau loss pada data validasi. Jika loss pada data validasi tidak menurun setelah 10 epoch, pelatihan akan dihentikan dan bobot model akan dikembalikan ke bobot terbaik yang telah disimpan sebelumnya. Proses pelatihan ini dilakukan secara verbose, yang berarti akan menampilkan progres pelatihan secara langsung selama proses berlangsung.

Gambar 5. Histogram BLEU pada data *training*Gambar 6. Histogram BLEU pada data *testing*

Hasil BLEU score pada Gambar 5. pada data pelatihan menunjukkan skor pada 1-grams, 1-2 grams, 1-3 grams dan 1-4 grams berturut turut adalah 0.995 , 0.985, 0.971 dan 0.9255. Model mesin penerjemah LSTM telah berhasil mencapai tingkat kesesuaian yang tinggi dalam memprediksi terjemahan yang mirip dengan referensi manusia. Skor yang tinggi untuk setiap n-gram, baik itu 1-grams, 1-2 grams, 1-3 grams, maupun 1-4 grams, menunjukkan bahwa model mampu menangkap struktur dan konteks kalimat dengan baik dalam data pelatihan.

Selanjutnya, BLEU score Gambar 6. pada data testing menunjukkan 0.994, 0.986, 0.971 dan 0.928. Model memiliki kemampuan yang baik dalam melakukan terjemahan pada data yang belum pernah dilihat sebelumnya. Perbedaan skor yang kecil antara data pelatihan dan data testing menunjukkan bahwa model tidak mengalami overfitting secara signifikan terhadap data pelatihan dan mampu melakukan generalisasi dengan baik pada data baru.

Tabel 9. Hasil terjemahan pada data *testing*. Kolom Halus merupakan bahasa sumber, kolom LOMA sebagai bahasa target dan kolom terjemahn merupakan hasil prediksi terjemahannya.

HALUS	LOMA	Terjemahan
teu damang ayeuna	teu cageur ayeuna	teu cageur ayeuna
rambut na oge	buuk na oge	buuk na oge
pami dicepeng murudul	lamun dicabak murudul	lamun dicabak murudul
anjeun janten istri kersa diwayuh	maneh jadi aweue daek	maneh jadi aweue daek
teu	dicandung teu	dicandung teu
hanteu sapertos di warung payun	teu siga di warung hareup	teu siga di warung hareup
arawis didinya mah	marahal didinya mah	marahal didinya mah
abdi mah teu hoyong meser	urang mah teu hayang meuli	urang mah teu hayang meuli
kadinya	kadinya	kadinya
saatos meser bumbu dapur	saenggeus meuli bumbu dapur	saenggeus meuli bumbu dapur

Pada Tabel 9. Hasil terjemahan pada model yang telah dibangun menghasilkan terjemahan yang baik. Model dengan arsitektur LSTM berhasil memecahkan kompleksitas *undak usuk basa* dalam Bahasa Sunda.

4. Kesimpulan

Penelitian ini menerapkan Neural Machine Translation (NMT) dengan algoritma Long Short Term Memory (LSTM) untuk meningkatkan terjemahan dari Bahasa Sunda Halus ke Bahasa Sunda Loma. Dengan penggunaan optimizer ADAM, model NMT berbasis LSTM dievaluasi dengan BLEU score, mencapai tingkat kesesuaian yang tinggi dengan referensi manusia pada data pelatihan dan testing. Skor BLEU yang tinggi menunjukkan kemampuan model dalam menangkap struktur dan konteks kalimat. Selain meningkatkan akurasi terjemahan, penelitian ini juga mengenalkan kontribusi teknologi NMT dalam melestarikan budaya bahasa daerah. Model yang dikembangkan diharapkan dapat menjadi dasar untuk pengembangan teknologi terjemahan otomatis dan pelestarian budaya bahasa daerah di Indonesia.

References

- [1] L. B. San and H. Sujaini, "Uji Nilai Akurasi pada Neural Machine Translation (NMT) Bahasa Indonesia ke Bahasa Tiochiu Pontianak dengan Mekanisme Attention," vol. 9, no. 3, pp. 362–370, 2023.
- [2] Ratnawati, R. Kusumah, and N. Cahyati, "Korelasi Peran Orang Tua terhadap Pemertahanan Bahasa Sunda sebagai Bahasa Ibu di daerah Kuningan," *Golden Age*, vol. 5, no. 02, pp. 474–481, 2021, [Online]. Available: <https://doi.org/10.29408/jga.v5i02.4387>
- [3] R. G. Guntara, A. Nuryadin, and B. Hartanto, "Pemanfaatan Google Speech to Text Untuk Aplikasi Pembelajaran Kamus Bahasa Sunda Pada Platform Mobile Android," *Justek J. Sains dan Teknol.*, vol. 4, no. 1, p. 10, 2021, doi: 10.31764/justek.v4i1.4455.
- [4] D. Mulyati, "Perancangan Aplikasi Pengenalan Bahasa Sunda Berbasis Android," *JIEET (Journal Inf. Eng. Educ. Technol.)*, vol. 2, no. 2, pp. 101–108, 2018.
- [5] N. R. D. C. Juliani *et al.*, "Pengenalan Undak Usuk Bahasa Sunda pada Anak Usia Dini di Desa Bakom," *Jassa J. Abdimas Sang Pencerah*, vol. 1, no. 1, pp. 14–17, 2023.
- [6] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection," *Inf. Process. Manag.*, vol. 59, no. 2, p. 102844, 2022, doi: <https://doi.org/10.1016/j.ipm.2021.102844>.
- [7] D. Anggara, N. Suarna, and Y. A. Wijaya, "ANALISA PERBANDINGAN PERFORMA OPTIMIZER ADAM, SGD, DAN RMSPROP PADA MODEL H5," *NERO (Networking Eng. Res. Oper.)*, vol. 8, no. 1, pp. 53–64, 2023.
- [8] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.
- [9] S. A. B. Andrabi and A. Wahid, "Machine Translation System Using Deep Learning for English to Urdu," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/7873012.