



Pengembangan Mesin Penerjemah Bahasa Indonesia ke Bahasa Daerah Kudus Menggunakan NMT Berbasis RNN-GRU

Anggay Bunga Permata^{1*}, Tri Listyorini², Endang Supriyati³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus, Kudus, Indonesia

^{1*} anggaybunga@gmail.com, ² trilistyorini@umk.ac.id, ³ endang.supriyati@umk.ac.id

^{*} anggaybunga@gmail.com

Abstrak-Bahasa Daerah Kudus merupakan identitas budaya yang penting, namun kini menghadapi ancaman kepunahan karena minimnya penggunaan oleh generasi muda dengan populernya bahasa gaul dan bahasa internasional serta kurangnya sumber daya teks digital. Teknologi mesin terjemah umu seperti Google Translate masih memiliki keterbatasan dalam memahami konteks budaya dan dialek lokal yang spesifik. Penelitian ini bertujuan untuk mengembangkan alat penerjemah otomatis dari Bahasa Indonesia ke Bahasa Daerah Kudus. Memanfaatkan teknologi *Neural Machine Translation* (NMT) berbasis RNN-GRU untuk mengatasi keterbatasan penerjemahan pada kategori *low resource language*. Metode yang digunakan adalah pendekatan kuantitatif eksperimental dengan menerapkan model NMT arsitektur *Recurrent Neural Network* (RNN) khususnya *Gated Recurrent Unit* (GRU). Data penelitian berupa korpus paralel dengan jumlah 1.031 pasang kalimat yang telah dibersihkan dan tokenisasi lalu dibagi menjadi tiga yaitu 824 untuk data *training*, 103 untuk data validasi dan 104 untuk data *testing*. Model dibangun menggunakan TensorFlow dilatih maksimum selama 106 *epoch* dengan *early stopping*, *optimizer* Adam dan *bactch* 64 pada GPU Colab. Hasil pengujian menunjukkan model yang dikembangkan mencapai skor BLEU 0,89% pada data pengujian, menunjukkan bahwa model masih berada pada tahap awal dan menghadapi tantangan besar dalam menangani data yang sangat terbatas (*low-resource*). Simpulan penelitian menegaskan bahwa model RNN-GRU memerlukan pengembangan lebih lanjut dengan memperluas jumlah korpus paralel dan mengeksplorasi lebih banyak arsitektur canggih untuk meningkatkan kekayaan kosakata dan akurasi.

Kata Kunci: Bahasa Daerah Kudus, NMT, RNN, GRU

Abstract-The Kudus regional language is an important cultural identity, but it is now facing extinction due to its lack of use by the younger generation, the popularity of slang and international languages, and the lack of digital text resources. General machine translation technologies such as Google translate still have limitations in understanding specific cultural contexts and lokal dialects. This research aims to develop an automatic translation tool from Indonesian to the Kudus regional language. Utilizing RNN-GRU-based Neural Machine Translation (NMT) technology to overcome translation limitations in the low resource language category. The method employed experimental quantitative approach by applying the NMT model of the Recurent Neural Network architecture, specifically Gated Recurent Unit (RNN-GRU). The dataset consists of a parallel corpus with a total 1,031 pairs of sentences that have been cleaned and tokenized, the divided into three parts 824 for training, 103 for validation, and 104 for testing. The model was built using Tensorflow and trained for a maximum of 106 epochs with early stopping, Adam optimizer, and bacth 64 on Colab GPU. The results demonstrate that the developed model achieved a BLEU score of 0.89% on the testing data, demonstrating the significant complexity and challenges of translating low-resource regional dialects. The findings suggest that the RNN-GRU model is further development is still needes by expanding the number of parallel corpora and exploring more state-of-the-art architectures to improve vocabulary richness and accuracy.

Keywords: Kudus Regional Languge, NMT, RNN, GRU

1. PENDAHULUAN

Bahasa memiliki peran yang penting dalam hal berkomunikasi baik dengan suara atau tulisan. Tidak hanya berfungsi sebagai alat pertukaran informasi, tapi juga sebagai identitas dan warisan kultural bagi penuturnya [1]. Sebagai negara kepulauan, Indonesia memiliki tanah luas dan pulau yang banyak [2]. Keberagaman suku dan geografis membuat setiap tempat memiliki kata dan cara ucap unik. Bedanya bahasa juga dapat disebabkan oleh faktor seperti tempat tinggal yang beda atau kelompok dalam masyarakat [3]. Begitupun dengan daerah Kudus yang berada di Jawa Tengah yang memiliki dialek khas yang bisa disebut Bahasa Daerah Kudus.

Namun, eksistensi warisan budaya ini kini dihasapkan pada tantangan besar seiring dengan kemajuan zaman yang serba cepat. Perkembangan teknologi yang telah merambah ke berbagai bidang memang





mempermudah aktivitas manusia [4]. Tapi di sisi lain, arus perkembangan zaman memperkuat dominasi Bahasa Indonesia sebagai bahasa nasional dan Bahasa Inggris sebagai bahasa internasional. Situasi ini secara tidak langsung mengancam keberlangsungan bahasa daerah kecil, termasuk Bahasa Daerah Kudus [1]. Saat ini, Bahasa Daerah Kudus menghadapi risiko kepunahan karena intensitas penggunaannya oleh generasi muda semakin menurun. Para pemuda cenderung lebih nyaman menggunakan bahasa nasional atau istilah asing dicampur dengan bahasa gaul karena dianggap modern dan mudah dipahami dalam pergaulan [5]. Fenomena tersebut menyebabkan terhambatnya proses penurunan bahasa ke generasi yang akan datang, yang tentu saja diperparah dengan minimnya ketersediaan informasi digital dalam Bahasa Daerah Kudus. Keterbatasan data teks digital ini menyebabkan Bahasa Daerah Kudus tergolong sebagai bahasa dengan *low-resource language*, yang memerlukan penanganan melalui teknologi agar dapat bertahan di zaman sekarang dan selanjutnya.

Upaya pelestarian dan revitalisasi bahasa daerah saat ini menjadi agenda yang sangat krusial, sehingga diperlukan sistem penerjemah otomatis yang mampu menjembatani bahasa nasional ke bahasa daerah guna mendukung pemahaman makna dan konteks komunikasi secara akurat [6]. Pemanfaatan teknologi *Machine Learning* (ML) telah terbukti mampu mempelajari pola data secara mandiri untuk memfasilitasi berbagai tugas manusia [7], kemudian hal ini dikembangkan lebih lanjut melalui *Deep Learning* (DL) dalam rangka mengolah pola data yang lebih kompleks dan rumit [8]. Di bidang kebahasaan, integrasi *Natural Language Processing* (NLP) dan *Machine Translation* (MT) menjadi pilar utama dalam pengembangan sistem penerjemah otomatis berbasis kecerdasan buatan (AI) [9]. Efektivitas AI dalam pengenalan pola bahasa telah ditunjukkan dalam berbagai penelitian, termasuk keberhasilan dalam pengenalan pola karakter yang menunjukkan tingkat akurasi tinggi [10].

Dalam perkembangannya, teknologi MT telah berevolusi dari *Statistical Machine Translation* (SMT) menuju *Neural Machine Translation* (NMT) yang mengadopsi prinsip kerja jaringan syaraf tiruan [11]. NMT menawarkan performa yang lebih unggul dibandingkan metode tradisional karena kemampuannya dalam memahami konteks bahasa secara lebih mendalam dan akurat [12]. Salah satu arsitektur NMT yang sangat efektif untuk menangani data urutan (sekuensial) adalah *Recurrent Neural Network* (RNN). Dalam penelitian ini, menggunakan varian *Gated Recurrent Unit* (GRU) karena kemampuannya mengatasi kelemahan mendasar pada RNN konvensional, khususnya dalam mengelola memori jangka panjang secara lebih efisien dan ringan saat proses komputasi [13]. Untuk mengevaluasi kualitas terjemahan yang dihasilkan, metrik *Bilingual Evaluation Understudy* (BLEU) digunakan sebagai standar pengukuran hasil yang diinginkan. Melalui RNN-GRU ini diharapkan menjadi solusi efektif dalam mengatasi keterbatasan korpus pada Bahasa Daerah Kudus dan membuat terjemahan lebih akurat [14].

Sejumlah penelitian sebelumnya oleh para peneliti yang berbeda telah meneliti penerjemah otomatis untuk bahasa daerah di Indonesia dari sisi pendekatan yang berbeda. Penelitian yang dilakukan oleh [15] berhasil membangun mesin penerjemah Bahasa Indonesia ke Bahasa Madura dialek Bangkalan dengan hasil skor BLEU sebesar 37,15%. Selanjutnya, [16] mengembangkan alat penerjemahan dari Bahasa Bangka ke Bahasa Indonesia berbasis web menggunakan model *Recurrent Neural Network* (RNN) yang menghasilkan skor BLEU sebesar 55,3%. Pendekatan RNN juga diterapkan oleh [17] untuk menerjemahkan Bahasa Indonesia ke Bahasa Melayu dialek Ketapang, yang mencapai performa skor BLEU sebesar 40,25%.

Meskipun Layanan penerjemah global seperti Google Translate telah banyak digunakan, penelitian oleh [18] mengungkapkan bahwa terdapat keterbatasan yang signifikan dalam memahami konteks budaya dan variasi dialek lokal. Seperti pada kasus Bahasa Batak. Google Translate dinilai cukup untuk menerjemahkan kata dasar, namun sering kali gagal dalam menangkap nuansa dialek lebih spesifik. Hal ini memertegas perlunya pengembangan model yang lebih terspesialisasi untuk bahasa daerah tertentu. Sejalan dengan kebutuhan tersebut, penelitian [19] membuktikan potensi model NMT berbasis arsitektur *Gated Recurrent Unit* (GRU) dalam menerjemahkan korpus paralel terbatas yang terdiri dari 1.302 kalimat, model ini menunjukkan kemampuan yang luar biasa untuk beradaptasi dengan karakter kompleks bahasa lokal.

Kesenjangan dalam penelitian saat ini adalah kurangnya studi mendalam yang menerapkan arsitektur RNN-GRU untuk Bahasa Daerah Kudus. Sebagian besar penelitian sebelumnya berfokus pada bahasa daerah dengan basis data yang lebih besar atau dialek yang berbeda secara struktural. Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut dengan mengimplementasikan NMT berbasis RNN-GRU pada Bahasa Kudus, yang diharapkan tidak hanya meningkatkan akurasi terjemahan melalui skor BLEU, tetapi juga mengatasi keterbatasan konteks lokal yang dimiliki oleh mesin penerjemah umum.

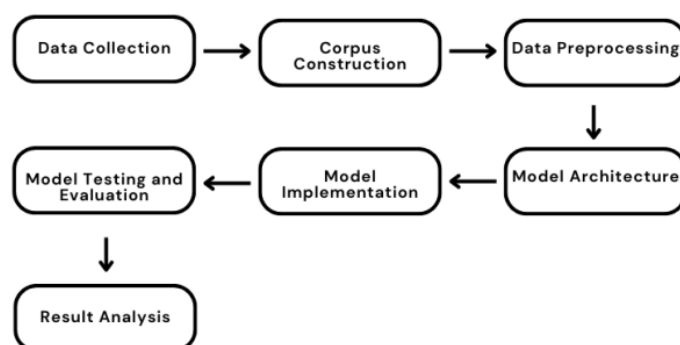
Kebaharuan dalam penelitian ini terletak pada pengembangan sistem Neural Machine Translation (NMT) pertama yang dikhususkan untuk pasangan Bahasa Indonesia ke Bahasa Daerah Kudus menggunakan arsitektur RNN-GRU. Berbeda dengan penelitian sebelumnya yang berfokus pada bahasa daerah dengan basis data besar, penelitian ini mengoptimasi penggunaan korpus (*low-resource language*) yang diintegrasikan dengan antarmuka interaktif berbasis Gradio.



Tujuan utama dari penelitian ini adalah membangun serta mengevaluasi model penerjemah otomatis yang mampu menghasilkan konversi teks secara akurat, yang diukur melalui metrik *Bilingual Evaluation Understudy* (BLEU). Hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam upaya pelestarian budaya digital, sekaligus menjadi fondasi bagi pengembangan model mesin penerjemah yang lebih kompleks dimasa depan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan pengembangan model *Neural Machine Translation* (NMT) dengan fokus implementasi Encoder-Decoder *Recurrent Neural Network* (RNN) khususnya *Gated Recurrent Unit* (GRU) untuk menerjemahkan Bahasa Indonesia ke Bahasa Daerah Kudus. Metode ini dikenal efektif dalam mengatasi masalah depensi jangka panjang pada data sekuens [20]. Model memanfaatkan data publik guna membuat korpus parallel untuk sumber data model. Menggunakan metrik *Bilingual Understudy Evaluation* (BLUE) untuk evaluasi kinerja model yang telah dibuat. Pelaksanaan penelitian ini ada beberapa langkah. Mulai dari pencarian data. Diakhiri dengan analisis hasil dan menarik kesimpulan. Alur metode penelitian dapat dilihat pada gambar 1.



Gambar 1. Alur metode penelitian

Beberapa tahapan metode penelitian ini meliputi:

2.1 Pengumpulan Data

Tahap pengumpulan data ini berfokus pada identifikasi dan pengumpulan data mentah. Data yang telah dikumpulkan digunakan sebagai dasar pembuatan korpus paralel. Proses ini mengambil data teks yang berbentuk frasa, kata atau kalimat Bahasa Daerah Kudus yang terdapat artinya dalam Bahasa Indonesia [19]. Semua data dikumpulkan menjadi satu dari berbagai platform daring. Tiga sumber data untuk pengumpulan data meliputi blog personal seperti bosokudus.blogspot, portal berita warga seperti kompasiana dan media sosial yaitu akun instagram bosos kudus.

2.2 Pembuatan Korpus

Korpus adalah kumpulan teks atau data lisan yang dibuat teratur dipilih dan dikumpulkan untuk menganalisis kata[2]. Pembuatan korpus paralel merupakan format yang sangat berguna untuk pembangunan mesin penerjemah [21]. Data mentah yang sudah terkumpul mulai dari frasa, kata dan kalimat diubah menjadi sepasang kalimat sederhana. Satu pasang kalimat sederhana terdiri dari kalimat Bahasa Daerah Kudus beserta artinya dalam Bahasa Indonesia. Kata yang memiliki arti serupa dan setara. Misalnya kata dalam Bahasa Indonesia dan Bahasa Daerah Kudus “Memasak-Kelan” diartikan dan disusun ulang sebagai sepasang kalimat sederhana “Memasak.” (dalam Bahasa Indonesia) dan “Kelan.” (dalam Bahasa Daerah Kudus). Dengan cara yang sama “Peci-Kethu” di ubah menjadi “Peci.”- “Kethu.”. Data yang didapat dari berbagai sumber membuat data sangat beragam dan tidak terorganisi karakternya jadi penyesuaian kaliman dilakukan secara manual. Memastikan setiap kalimat asli (Bahasa Indonesia) memiliki arti dan makna yang tepat sesuai dengan bahasa tujuan (Bahasa Daerah Kudus). Setelah disusun menjadi sebuah korpus terdapat 1044 baris pasang kalimat sederhana. Contoh korpus dapat dilihat pada tabel 1.

Tabel 1. Korpus Paralel

Bahasa Indonesia	Bahasa Kudsan
Jauh.	Adoh.
Jauh sekali.	Uaduh.
Tidak tahu.	Amboh.
Tidak mengerti.	Emboh.
Dengan.	Ambi.
Rusak.	Amoh.
Tawar.	Anyep.
Akan.	Ape.
Jatuh.	Ceblok.
Gigit.	Cokot.

2.3 Pra processing data

Korpus yang sudah jadi perlu diproses sebelum dijalankan dimodel. Pra pemrosesan data ini dilakukan untuk mempersiapkan korpus dengan cara membersihkan data, tokenisasi sampai ke pembagian data. Korpus disiapkan agar sesuai untuk di inputkan ke mesin penerjemah [21]. Serangkaian pra pemrosesan meliputi:

- Pembersihan data: Proses penghapusan data mencakup penghapusan data yang sama atau bisa disebut data duplikat dan menghapus baris yang kosong dalam korpus yang sudah di input. proses ini bertujuan meningkatkan kualitas representasi data dengan menghapus elemen yang tidak relevan (*noise*) dari korpus yang jadi [22]. Korpus parallel dengan 1044 baris data setelah dilakukan pembersihan menjadi 1031. Oleh karena itu, data bersih yang siap digunakan untuk model NMT sebanyak 1031 baris kalimat sederhana.
- Pembagian Data: Korpus paralel yang sudah bersih akan dibagi tiga. Ada data untuk latihan, data untuk validasi dan data untuk pengujian [23]. Bagian ini ikut pembagian yang biasa dipakai. Data latihan dapat 80%. Data validasi dapat 10%. Data tes dapat 10% [23]. Total data dari presentase tersebut menjadi 824 pasang kalimat untuk data training, Jumlah data untuk latihan ada 824 pasang kalimat. Untuk validasi ada 103 pasang kalimat biasa. Untuk tes ada 104 pasang kalimat biasa. Semua data sudah dibagi sesuai rasio itu.
- Tokenisasi: Masing – masing kalimat input (Bahasa Indonesia) dan kalimat target (Bahasa Daerah Kudus) melalui tahap tokenisasi ini. Mengubah kalimat yang tadinya berupa teks diubah menjadi urutan angka [24]. Beberapa informasi setelah tokenisasi berhasil dapat dilihat jumlah kata unik input sebanyak 1116 kata, jumlah kata unik target sebanyak 1094, panjang maksimum input 18 dan panjang maksimum target 9.

2.4 Pembangunan model

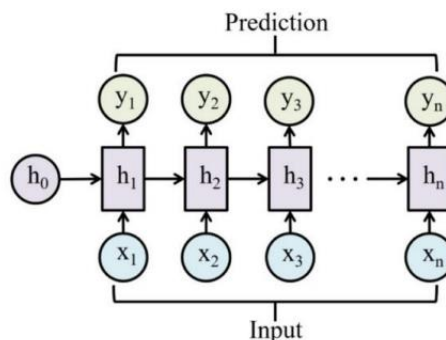
Pembangunan model NMT telah diimplementasikan dan dikonfigurasi sesuai kebutuhan mesin penerjemah. Struktur Encoder-Decoder merupakan dasar dari NMT dengan penerapan RNN-GRU ini digunakan untuk membuat model penerjemah. Struktur ini dipilih karena kemampuannya yang dikenal efektif dalam mengatasi tantangan depensi jangka panjang pada data sekuens [20]. RNN dirancang untuk memproses data berurutan dengan mempertahankan status tersembunyi. Ada lapisan yang tersembunyi dan lapisan yang keluar. RNN punya sambung yang berulang. Ini berbeda dari jaringan lain yang keluar. Hal ini buat data bisa jalan lagi di dalam jaringan. RNN ambil vector input x_t . RNN juga ubah vector tersembunyi h_t . RNN pakai rumus berikut ini:

$$h_t = \hat{y}h(LxTxT + Ph\check{Y}1 + b_h), \quad (1)$$

Dimana W_{xh} adalah matriks bobot yang terletak antara lapisan input dan lapisan tersembunyi, W_{hh} adalah bobot matriks koneksi berulang, b_h adalah vector bias, dan $\hat{y}h$ fungsi aktivasi yang biasanya merupakan fungsi anh atau unit linier yang diperbaiki, disebut h. Selanjutnya, persamaan digunakan untuk menghitung keluaran pada setiap langkah waktu t dihitung dengan persamaan:

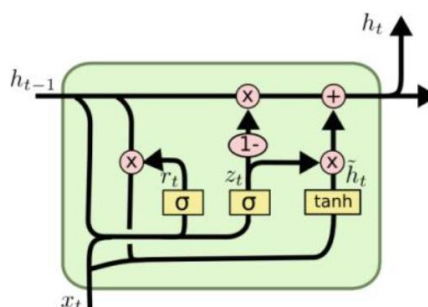
$$y_t = \hat{y}y(W_{hy}h_t + b_y), \quad (2)$$

W_{hy} merupakan matriks bobot antara lapisan tersembunyi dan lapisan keluaran, b_y adalah vector bias, serta $\hat{y}y$ adalah fungsi aktivasi yang digunakan pada lapisan luaran. Arsitektur dasar RNN ditunjukkan pada gambar 2.



Gambar 2. Arsitektur RNN dasar [25]

Dalam ranah kerja pada data teks, RNN mulai pakai kata yang telah jadi angka. Ini artinya tiap kata diubah ke bentuk angka dulu. Lalu, angka tiap kata itu dikumpul dan disusun jadi baris dalam tabel. Tabel ini lalu masuk ke dalam RNN. Dalam proses ini, model tahu apa arti kata. Model juga tahu kaitan kata satu sama lain [26]. Sedangkan GRU sendiri memiliki dua pintu. Pintu itu *reset* (r_t) dan *update* (z_t). Secara spesifik, *reset gate* pada GRU bertanggung jawab untuk menentukan seberapa banyak informasi yang tidak relevan dari masa lalu yang perlu diabaikan atau dilupakan. Sementara, *update* berfungsi untuk mengontrol seberapa banyak memori dari informasi sebelumnya yang perlu dipertahankan dan diteruskan ke angka waktu berikutnya. Arsitektur sel GRU dapat dilihat pada gambar 3.



Gambar 3. Arsitektur GRU

Model GRU punya tiga lapis. lapis *input*, lapis *output* dan lapis lain. Data masuk ke GRU pakai jenis waktu. Sebelum dipakai, data ini diproses dulu. Cara kerja di sel GRU yaitu *Reset gate* (r_t) dipilih. *Gate* ini atur informasi lama. Informasi yang sudah tidak perlu bisa dibuang. Perhitungan ini melibatkan *input* saat ini (x_t) dan *hidden state* dari waktu sebelumnya (h_{t-1}) dikalikan dengan bobot (w_r).

$$r_t = \sigma (W_r \cdot [h_{t-1}, x_t]) \tag{3}$$

Penentuan *update gate* (z_t) untuk mengukur seberapa banyak informasi dari masa lalu yang akan diteruskan ke masa depan. *Gate* ini penting untuk mencegah *vanishing gradient* dengan memungkinkan model menyalin informasi lama secara utuh. Perhitungan melibatkan *input* saat ini (x_t) dan *hidden state* dari waktu sebelumnya (h_{t-1}), dikalikan dengan bobot (w_z).

$$z_t = \sigma (W_z \cdot [h_{t-1}, x_t]) \tag{4}$$

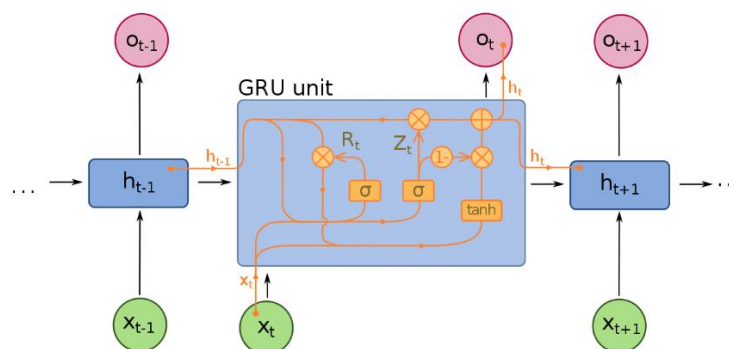
Penetapan *candidate hidden state* (\hat{h}_t), untuk menghasilkan keluaran sementara pada waktu t menggunakan *tanh layer* dan *reset gate* (r_t), untuk mengembalikan informasi relevan dari masa lalu.

$$\hat{h}_t = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t]) \tag{5}$$

Penetapan output standar unit GRU (h_t), untuk menyimpan informasi dari unit saat ini, dengan memadukan output dari *update gate* (z_t), *candidate hidden state* (\hat{h}_t), dan *hidden state* dari waktu sebelumnya (h_{t-1}) [27].

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h_t \quad [27]. \quad (6)$$

Encoder sendiri bertanggung jawab untuk memproses sekuens input (Bahasa Indonesia) dan mengkonversinya menjadi representasi vector kontekstual. Sedangkan Decoder untuk menghasilkan sekuens target (Bahasa Daerah Kudus) secara *autoregressive* memanfaatkan hasil representasi vector dari encoder untuk prediksi kata berikutnya secara bertahap. Struktur ini tersusun dari beberapa lapisan utama yaitu lapisan input, embedding, GRU encoder, GRU decoder dan lapisan dense (output) untuk memetakan hasil ke kosakata target. Gambar arsitektur RNN-GRU dapat dilihat pada gambar 4.



Gambar 4. Arsitektur RNN-GRU

2.5 Implementasi model

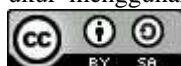
Proses implementasi model melibatkan beberapa proses pelatihan model dan penyesuaian parameter untuk mencapai kinerja model yang optimal. Model dilatih menggunakan data *training*. Meminimalisir fungsi *loss* yang mengukur ketidaksesuaian antara prediksi terjemahan model dan terjemahan acuan. Kinerja model tetap dipantau dengan data validasi selama proses pelatihan berlangsung. Hal ini berfungsi untuk mendeteksi *overfitting* dan menentukan titik henti peletihan yang efektif [28]. Eksperimen dilakukan untuk mengatur *hyperparameter* guna memastikan model beroperasi pada kinerja maksimal sesuai prinsip *deep learning* [29]. Total 1,636,422 parameter (6.24 MB) bersifat *trainable*, akan diperbarui saat proses pelatihan berjalan. Lebih lengkapnya dapat dilihat gambar 5.

Layer (type)	Output Shape	Param #	Connected to
encoder_input (InputLayer)	(None, 18)	0	-
decoder_input (InputLayer)	(None, 8)	0	-
embedding (Embedding)	(None, 18, 256)	285,696	encoder_input[0]...
not_equal (NotEqual)	(None, 18)	0	encoder_input[0]...
embedding_1 (Embedding)	(None, 8, 256)	288,064	decoder_input[0]...
gru (GRU)	[(None, 256), (None, 256)]	394,752	embedding[0][*], not_equal[0][*]
gru_1 (GRU)	[(None, 8, 256), (None, 256)]	394,752	embedding_1[0][*], gru[0][*]
dense (Dense)	(None, 8, 1094)	281,158	gru_1[0][*]

Gambar 5. Functional Model

2.6 Pengujian dan evaluasi model

Tahap pengujian dan evaluasi model sangat penting untuk mengukur efektifitas dan kualitas model penerjemahan yang dikembangkan [30]. Setelah model selesai dilatih, model diuji secara mandiri menggunakan data *testing*. Selama pelatihan, data *testing* ini belum pernah diakses oleh model. Kualitas model juga akan diukur menggunakan metrik standar yaitu BLEU Score (*Bilingual Evaluation Understudy*) [14]. Dengan cara



menghitung presisi n -gram dari terjemahan yang dihasilkan model dibandingkan dengan satu atau lebih terjemahan.

Secara matematis, perhitungan *brevity penalty* (PB) dan skor BLEU dapat dijelaskan sebagai berikut:

Rumus Brevity Penalty (PB):

$$P_B = \begin{cases} 1, & c > r \\ e^{(1-\frac{r}{c})}, & c \leq r \end{cases} \quad (7)$$

Rumus Skor BLEU:

$$BLEU = P_B \exp \left(\sum_n^N w_n \log P_n \right) \quad (8)$$

Dimana:

- r adalah panjang korpus referensi.
- c adalah panjang kandidat terjemahan (hasil mesin).
- w_n adalah bobot positif yang dijumlahkan menjadi satu, merepresentasikan kontribusi setiap n -gram (misalnya, w_1 untuk *unigram*, w_2 untuk *bigram* dan seterusnya)
- p_n adalah kalkulasi presisi n -gram pada n maksimum, mengukur jumlah n -gram yang cocok antara terjemahan kandidat dan referensi.

Dengan demikian, BLEU menawarkan cara yang objektif untuk mengukur kualitas terjemahan mesin dengan membandingkan kesamaan tekstual antara keluaran mesin dan terjemahan referensi manusia, sambil mempertimbangkan baik presisi maupun panjang terjemahan [6].

2.7 Analisis Hasil

Tahap terakhir dari penelitian ini adalah analisis hasil dengan merumuskan kesimpulan penelitian dari hasil evaluasi akan diinterpretasikan. Skor BLEU menjadi salah satu faktor untuk mengukur keberhasilan model dalam mencapai tujuan penelitian. Analisis keefektifan model dalam menerjemahkan frasa, kata atau kalimat yang memerlukan pengembangan lebih lanjut. Dilakukan juga perumusan kesimpulan penelitian dan identifikasi potensi arah penelitian

3. HASIL DAN PEMBAHASAN

3.1 Uraian Tahapan Model

Berdasarkan hasil eksperimen yang telah dipaparkan, penelitian ini membuktikan bahwa implementasi arsitektur *Encoder-Decoder* berbasis RNN-GRU telah berhasil membangun sistem penerjemah Bahasa Indonesia ke Bahasa Daerah Kudus.

3.2.1 Analisis korpus dan pra-pemrosesan

Penelitian ini berhasil dalam pengumpulan kalimat bahasa Indonesia ke Bahasa Daerah Kudus. Data berupa pasangan kalimat sederhana yang digunakan untuk pembuatan korpus. Korpus yang berhasil dibuat terdiri dari 1044 pasang kalimat sederhana. Setelah di bersihkan pada tahap pra pemrosesan data yang siap pakai menjadi 1031 baris [31]. Data dibagi menjadi tiga yaitu data *training* sejumlah 824 baris, data validasi sejumlah 103 kalimat dan data *testing* sejumlah 104 baris. Hasil dari pembersihan dapat dilihat pada gambar 6.

```
# Cek ulang hasilnya
print("Jumlah data setelah dibersihkan:", len(df))
print(df.sample(3))

Jumlah data setelah dibersihkan: 1031
Bahasa Indonesia                                     Boso Kudus
755 Daripada dilempari. <start> Timbang diantemi. <end>
2 Tidak Tahu. <start> Amboh. <end>
822 Makan banyak. <start> Mangan akeh. <end>
```

Gambar 6. Jumlah Data Bersih

Tahap pembersihan data dalam pengembangan korpus paralel ini merupakan fondasi utama untuk memastikan model *Neural Machine Translation* (NMT) tidak mempelajari pola yang menyesatkan. Proses penghapusan data

mencakup identifikasi dan eliminasi data yang bersifat redundan atau duplikat, di mana kalimat yang sama muncul lebih dari satu kali dalam dataset. Keberadaan data duplikat ini sangat berisiko menciptakan bias frekuensi yang tidak alami, yang pada akhirnya dapat memaksa model untuk melakukan *overfitting* pada struktur kalimat tertentu saja. Selain itu, penghapusan baris kosong atau entri yang tidak lengkap juga dilakukan secara teliti untuk menghindari kegagalan komputasi saat proses *tokenization* dan *embedding*.

Secara teknis, proses ini bertujuan untuk meningkatkan kualitas representasi data dengan cara membuang seluruh elemen yang tidak relevan, atau yang secara luas dalam disiplin *Natural Language Processing* disebut sebagai *noise* [22]. *Noise* ini dapat berupa simbol non-alfanumerik yang tidak bermakna, spasi berlebih, hingga inkonsistensi penulisan karakter yang dapat mengaburkan makna semantik antarkata. Korpus paralel yang pada awalnya berjumlah 1.044 pasangan kalimat, setelah melalui filter pembersihan yang ketat, menyisakan 1.031 baris data yang benar-benar bersih dan siap pakai (*ready-to-use*). Penyusutan sebanyak 13 baris data ini, meskipun terlihat kecil secara kuantitas, memiliki dampak signifikan terhadap efisiensi pelatihan model. Dengan hanya menyisakan 1.031 baris kalimat sederhana yang berkualitas tinggi, model diharapkan dapat menangkap relasi sintaksis antara Bahasa Indonesia dan Bahasa Daerah Kudus secara lebih akurat tanpa terdistraksi oleh data sampah yang dapat meningkatkan nilai *error* atau *loss* pada tahap evaluasi nantinya. Setelah itu proses pembagian data, hasil dari pembagian data dapat dilihat pada gambar 7.

```
print(" Data berhasil dibagi:")
print(f"Train set : {len(input_train)} kalimat")
print(f"Validation : {len(input_val)} kalimat")
print(f"Test set : {len(input_test)} kalimat")

# Contoh data
print("\nContoh pasangan data:")
for i in range(2):
    print(f"- {input_train[i]} + {target_train[i]}")

Data berhasil dibagi:
Train set : 824 kalimat
Validation : 103 kalimat
Test set : 104 kalimat

Contoh pasangan data:
- Makanya sampai sekarang. + <start> Mulane nganti sak :
- Saya tendang pakai tumit kepalamu nanti. + <start> Tal
```

Gambar 7. Pembagian Data

Setelah melalui tahapan pembersihan yang ketat, korpus paralel Bahasa Indonesia dan Bahasa Daerah Kudus yang telah dinyatakan bersih (*clean corpus*) kemudian masuk ke tahap distribusi data. Secara metodologis, korpus ini dibagi menjadi tiga subset utama yang memiliki peran berbeda dalam siklus pengembangan model, yaitu data pelatihan (*training set*), data validasi (*validation set*), dan data pengujian (*testing set*) [23]. Pembagian ini mengikuti standar konvensi yang umum digunakan dalam penelitian pembelajaran mesin (*machine learning*), di mana mayoritas data dialokasikan untuk proses pembelajaran guna memastikan model mampu menangkap kompleksitas struktur linguistik yang ada.

Rasio pembagian yang diterapkan adalah 80% untuk data pelatihan, 10% untuk data validasi, dan 10% untuk data pengujian [23]. Dengan total korpus bersih sebanyak 1.031 baris, implementasi rasio tersebut menghasilkan distribusi yang sistematis: sebanyak 824 pasang kalimat didedikasikan sebagai data latihan yang berfungsi untuk memperbarui bobot (*weights*) pada jaringan saraf RNN-GRU. Sementara itu, 103 pasang kalimat dialokasikan untuk data validasi yang berperan vital dalam memantau performa model selama proses pelatihan, sekaligus menjadi acuan bagi fitur *early stopping* guna menghindari *overfitting*. Terakhir, 104 pasang kalimat ditetapkan sebagai data pengujian (*testing*) untuk mengevaluasi kemampuan generalisasi model terhadap data baru yang belum pernah ditemui sebelumnya. Pembagian yang proporsional ini sangat krusial, terutama pada bahasa daerah dengan sumber daya rendah (*low-resource*), agar setiap segmen data memberikan kontribusi maksimal dalam membentuk mesin penerjemah yang stabil dan akurat.

Setelah tahapan pra-pemrosesan dan pembagian data selesai dilakukan, langkah krusial berikutnya adalah mentransformasikan data tekstual menjadi format numerik yang dapat dipahami oleh arsitektur *Neural Machine Translation* (NMT). Proses ini diawali dengan tokenisasi, yaitu prosedur pemecahan rangkaian kalimat menjadi satuan unit terkecil yang disebut token. Dalam penelitian ini, sistem tokenisasi tidak hanya memecah kata, tetapi juga menyisipkan token penanda khusus, yaitu *<start>* dan *<end>*, pada setiap kalimat target (Bahasa Daerah Kudus). Token *<start>* berfungsi sebagai pemicu (*trigger*) bagi *decoder* untuk mulai membangkitkan kata pertama, sedangkan token *<end>* berfungsi sebagai sinyal penghenti agar model berhenti menghasilkan kata setelah struktur kalimat dianggap lengkap secara semantik.

Berdasarkan hasil ekstraksi pada korpus yang telah dibangun, proses ini berhasil mengidentifikasi sebanyak 1.116 kata unik untuk Bahasa Indonesia dan 1.094 kata unik untuk Bahasa Daerah Kudus [32]. Perbedaan jumlah kosa kata ini menunjukkan bahwa meskipun kedua bahasa memiliki akar linguistik yang serumpun, terdapat variasi dalam kekayaan leksikal pada dataset yang digunakan. Selain identifikasi kata unik, analisis terhadap panjang sekuens menjadi aspek teknis yang sangat fundamental. Hasil observasi menunjukkan adanya perbedaan signifikan pada panjang urutan maksimum, di mana Bahasa Indonesia memiliki panjang maksimum 18 token, sedangkan Bahasa Daerah Kudus memiliki 9 token [32]. Perlu dicatat bahwa perhitungan panjang token pada Bahasa Kudus ini sudah mencakup penambahan token spesial <start> dan <end>.

Untuk menyeragamkan dimensi input agar dapat diproses secara *batch* dalam jaringan saraf, diterapkan teknik *padding*. Dengan *padding*, setiap kalimat yang memiliki panjang di bawah ambang batas maksimum akan diisi dengan token khusus (seperti <PAD> atau nilai 0) hingga mencapai panjang standar. Hal ini memastikan bahwa matriks input yang masuk ke dalam lapisan *Embedding* memiliki bentuk yang konsisten. Tanpa adanya sinkronisasi panjang sekuens dan penggunaan token penanda batas, model RNN-GRU akan kesulitan menentukan kapan sebuah kalimat terjemahan harus dimulai dan diakhiri, yang pada akhirnya dapat menyebabkan kegagalan dalam pembentukan struktur kalimat yang koheren. Detail distribusi dan hasil akhir dari proses tokenisasi ini dapat diamati secara lebih komprehensif pada Gambar 8.

```
print("✅ Tokenisasi selesai!")
print(f"Jumlah kata unik (input) : {num_encoder_tokens}")
print(f"Jumlah kata unik (target) : {num_decoder_tokens}")
print(f"Panjang maksimum (input) : {max_len_input}")
print(f"Panjang maksimum (target) : {max_len_target}")

✅ Tokenisasi selesai!
Jumlah kata unik (input) : 1116
Jumlah kata unik (target) : 1094
Panjang maksimum (input) : 18
Panjang maksimum (target) : 9
```

Gambar 8. Tokenisasi

3.2.2 Implementasi arsitektur dan pelatihan model

Mesin penerjemah Bahasa Indonesia ke Bahasa Daerah Kudus berhasil dibangun dan di implementasi menggunakan *Encoder-Decoder* berbasis RNN-GRU. *Encoder* berfungsi untuk menginput melalui lapisan *embedding* sedangkan, *Decoder* secara *autoregressive* memprediksi output menggunakan GRU [33].

Model ini efektif dalam mengatasi depensi jangka panjang pada data sekuens terbatas [20]. Penerapan Jaringan Saraf Tiruan sebelumnya telah dirintis melalui sistem pengenalan Bahasa Isyarat Indonesia (SIBI) yang menggunakan algoritma *backpropagation* untuk klasifikasi pola gerak [34].

Pelatihan model dilakukan selama 106 *epoch* menggunakan optimizer Adan dengan *batch size* 64. Penggunaan *early stopping* untuk mencegah terjadinya *overfitting* pada model. Hasilnya menunjukkan *training loss* sebesar 0,0333, *validation loss* 6,4048 dan akurasi pelatihan 0,4879. Memperlihatkan penurunan tajam pada *ttraining loss*. *Validation loss* relative stabil pada *epoch* ke 20. Hal ini memperkirakan keberhasilan regulasi dengan terbatasnya data yang ada [35].

3.2 Hasil Pengujian Model

Evaluasi performa menggunakan metrik *Bilingual Evaluation Understudy* (BLEU). Pengujian dilakukan secara objektif terhadap 104 pasang kalimat testing yang sepenuhnya terisolasi dari proses pelatihan, guna mengukur kemampuan generalisasi model terhadap data baru. Diperoleh skor rata-rata BLEU sebesar 0,89%. Secara teknis, skor ini merepresentasikan tingkat kecocokan n-gram antara urutan kata yang dihasilkan oleh decoder RNN-GRU dengan kalimat referensi asli yang disusun oleh pakar bahasa atau penutur asli dialek Kudus.

Perolehan skor 0,89% memberikan gambaran empiris bahwa arsitektur RNN-GRU menghadapi tantangan besar dalam melakukan pemetaan fitur linguistik pada kondisi *extremely low-resource language*. Dengan ketersediaan korpus yang hanya berjumlah 1.031 baris data bersih, model memiliki keterbatasan dalam membangun ruang *embedding* yang kaya akan variasi kosakata dan struktur sintaksis dialek Kudus yang kompleks. Fenomena ini mengonfirmasi bahwa model *deep learning* berbasis *recurrent* sangat bergantung pada skala data untuk mencapai tingkat akurasi fungsional. Namun, dalam konteks penelitian bahasa daerah kudus yang belum memiliki dokumentasi digital sebelumnya, angka ini bukanlah sekadar representasi kegagalan sistem, melainkan sebuah skor acuan (*baseline*) yang sangat berharga.

Hasil penelitian ini berhasil menjawab kesenjangan teknologi Natural Language Processing (NLP) pada bahasa daerah. Implikasi dari temuan ini sangat signifikan, penelitian ini telah meletakkan fondasi teknologi digital pertama untuk Bahasa Daerah Kudus yang terstruktur. Meskipun cakupan kosakata masih terbatas, keberhasilan pengintegrasian arsitektur Encoder-Decoder ini membuktikan bahwa alur translasi otomatis dialek lokal dapat diwujudkan. Keterbatasan pada aspek akurasi ini menjadi titik tolak krusial bagi peneliti selanjutnya untuk melakukan ekspansi korpus paralel secara masif atau mengeksplorasi teknik transfer learning guna mengatasi kendala data yang terbatas, demi menjaga eksistensi Bahasa Kudus di era kecerdasan buatan.

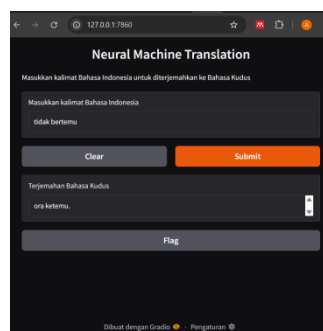
3.3 Implementasi Prototipe Antarmuka Pengguna

Tahap terakhir yaitu pengembangannya menjadi sebuah mesin translate sendan antarmuka pengguna. Dibuat menggunakan *Gradio* tampilan responsive berhasil diimplementasikan. Tampilan antarmuka mesin terjemah ini dapat dilihat pada gambar 9.

Tahap akhir dari rangkaian penelitian ini adalah transformasi model saraf tiruan yang abstrak menjadi sebuah aplikasi praktis berupa mesin penerjemah dengan antarmuka pengguna (User Interface). Pengembangan prototipe ini direalisasikan menggunakan library *Gradio*, sebuah kerangka kerja berbasis Python yang dirancang khusus untuk mendemonstrasikan model *machine learning* secara interaktif. Pemilihan *Gradio* didasarkan pada efisiensinya dalam mengintegrasikan fungsi inferensi model RNN-GRU ke dalam elemen visual web tanpa memerlukan pengembangan backend yang kompleks, namun tetap menghasilkan tampilan yang profesional dan responsif.

Secara teknis, prototipe ini bekerja dengan menghubungkan skrip prediksi model .h5 yang telah dilatih dengan komponen input-output pada antarmuka. Ketika pengguna memasukkan rangkaian kalimat dalam Bahasa Indonesia pada kolom yang tersedia, sistem secara otomatis menjalankan fungsi pra-pemrosesan di latar belakang-termasuk tokenisasi dan padding-sebelum meneruskannya ke arsitektur Encoder-Decoder. Hasil prediksi dari unit GRU kemudian dikonversi kembali dari bentuk vektor numerik menjadi teks manusia dalam Bahasa Daerah Kudus, lalu ditampilkan secara real-time pada panel output.

Implementasi antarmuka ini memiliki desain yang responsif, yang berarti tata letak elemennya dapat menyesuaikan secara otomatis dengan berbagai ukuran layar, baik diakses melalui perangkat desktop. Keberhasilan pembangunan antarmuka ini membuktikan bahwa teknologi *Natural Language Processing* (NLP) untuk bahasa daerah yang minim sumber daya (*low-resource*) dapat diakses secara luas. Prototipe ini tidak hanya berfungsi sebagai alat uji coba bagi peneliti, tetapi juga menjadi fondasi awal bagi pengembangan kamus digital interaktif yang dapat membantu pelestarian dialek lokal di era digital. Detail tampilan visual dari mesin penerjemah ini, yang mencakup area input teks dan hasil terjemahan, dapat diamati secara komprehensif pada Gambar 9.



Gambar 9. Interface Model

3.4 Pembahasan dan Perbandingan

Hasil evaluasi skor BLEU sebesar 0,89 pada penelitian ini memberikan perspektif baru dalam pengembangan Neural Machine Translation (NMT) untuk dialek lokal yang spesifik seperti Bahasa Daerah Kudus. Meskipun secara numerik skor ini berada pada kategori baseline, pencapaian ini harus dianalisis dalam konteks keterbatasan sumber daya (*low-resource language*). Perolehan skor tersebut membuktikan bahwa arsitektur Encoder-Decoder berbasis RNN-GRU mampu melakukan pemetaan semantik dasar, meskipun menghadapi tantangan besar pada variasi bahasa daerah yang belum terstandarisasi dalam korpus digital.



Jika dibandingkan dengan penelitian sejenis, terdapat beberapa poin penting yang memperkuat posisi penelitian ini. Sebagai contoh, penelitian NMT pada Bahasa Madura [15] atau Bahasa Sunda [17] yang menggunakan arsitektur serupa seringkali mencapai skor yang lebih tinggi karena didukung oleh volume korpus paralel yang mencapai 5000 baris data. GRU memiliki parameter yang lebih sedikit namun tetap mampu menjaga ketergantungan sekuensial (sequential dependencies), sehingga sangat cocok diimplementasikan pada dataset berskala kecil tanpa membebani sumber daya komputasi secara berlebihan.

Secara keseluruhan, diskusi ini menegaskan bahwa model RNN-GRU yang dikembangkan telah memenuhi fungsinya sebagai prototipe penerjemah fungsional. Implikasi dari perbandingan ini menyarankan bahwa untuk meningkatkan skor BLEU di masa mendatang, diperlukan memperbanyak data dengan strategi augmentasi data.

4. KESIMPULAN

Penelitian ini telah berhasil mengembangkan sebuah sistem mesin penerjemah otomatis dari Bahasa Indonesia ke Bahasa Daerah Kudus dengan menerapkan arsitektur *Neural Machine Translation* (NMT) berbasis *Recurrent Neural Network-Gated Recurrent Unit* (RNN-GRU). Melalui tahapan eksperimental, penelitian ini menggunakan korpus paralel sebanyak 1.031 pasang kalimat yang dikurasi secara manual dari berbagai sumber digital sebagai dataset primer. Hasil evaluasi menggunakan metrik *Bilingual Evaluation Understudy* (BLEU) menunjukkan perolehan skor rata-rata sebesar 0,89%. Angka ini menjadi nilai acuan (*baseline*) pertama bagi pengembangan teknologi bahasa untuk dialek Kudus, yang sekaligus membuktikan bahwa model *deep learning* dapat diimplementasikan pada kategori *low-resource language* meskipun dengan keterbatasan jumlah data yang signifikan. Secara teknis, penggunaan unit GRU terbukti efisien dalam memproses sekuens bahasa daerah, namun performa akhir menunjukkan ketergantungan yang sangat tinggi terhadap skala dan variasi kosakata dalam korpus. Kontribusi utama dari penelitian ini adalah terciptanya dokumentasi digital awal bagi Bahasa Daerah Kudus yang dapat menjadi fondasi bagi pelestarian budaya lokal berbasis teknologi informasi. Untuk pengembangan di masa mendatang, sangat direkomendasikan untuk melakukan ekspansi dataset secara masif serta mengeksplorasi penggunaan mekanisme attention atau arsitektur Transformer untuk melihat perbandingan akurasi pada model yang digunakan. Penelitian ini menegaskan bahwa dokumentasi bahasa melalui mesin penerjemah merupakan langkah strategis dalam menjaga eksistensi dialek lokal di era digital, namun memerlukan kolaborasi berkelanjutan dalam pengayaan data linguistik agar mencapai tingkat akurasi fungsional yang lebih optimal.

REFERENSI

- [1] W. Murtafi'ah and S. K. Bariah, "Peran Bahasa Daerah dalam Identitas Nasional : Masihkah Relevan di Era Globalisasi ?," *JoALLE J. Adv. Lang. Lit. Educ.*, vol. 1, no. 1, pp. 12–16, 2025, [Online]. Available: <https://ejournal.dwipantara.org/index.php/joalle>
- [2] D. Soyusiawaty, A. Fadlil, and Sunardi, "Pengembangan Korpus Bahasa Minang pada Spell Error Corpus for Minang Language (SPEML)," *J. Nas. Teknol. dan Sist. Inform.*, vol. 01, pp. 17–26, 2025.
- [3] A. Zamhari, Aina, Fitriani, Waliya, S. Ari, and D. Hadisatya, "BAHASA SEBAGAI PENDIDIKAN BUDAYA DAN KARAKTER BANGSA," *EDUKREATIF J. Kreat. DALAM Pendidik.*, vol. 6, no. 1, pp. 1–14, 2025, [Online]. Available: <https://ijurnal.com/1/index.php/jkp>
- [4] D. E. Saputra, E. Supriyati, and T. Listyorini, "DIGITALISASI BATIK KUDUS DALAM PELESTARIAN BUDAYA DAERAH," *JTe*, vol. 9, pp. 178–184, 2024, doi: <https://doi.org/10.52561/teknika.v9i2.395>.
- [5] I. P. S. R, N. N. Insani, and M. R. Ridha, "Ancaman Pergeseran Bahasa Daerah Dan Dampaknya Terhadap Keberlanjutan Warisan Budaya Di Era Global," *J. Penelit. Nusant.*, vol. 1, pp. 91–96, 2025, doi: <https://doi.org/10.59435/menulis.v1i5.236>.
- [6] Y. Fauziyah, R. Ilyas, and F. Kasyidi, "Mesin Penterjemah Bahasa Indonesia-Bahasa Sunda Menggunakan Recurrent Neural Networks," *J. Teknoinfo*, vol. 16, no. 2, p. 313, 2022, doi: [10.33365/jti.v16i2.1930](https://doi.org/10.33365/jti.v16i2.1930).
- [7] E. Z. Astuti, C. A. Sari, E. H. Rachmawanto, and R. R. Ali, "Comparative Study of Machine Learning Algorithms for Performing Ham or Spam Classification in SMS," *Sci. J. Informatics*, vol. 11, no. 1, pp. 177–186, 2024, doi: [10.15294/sji.v11i1.47364](https://doi.org/10.15294/sji.v11i1.47364).
- [8] W. Andriani, S. Anandianskha, and A. A. Murtopo, "Optimization Selection on Deep Learning Algorithm for Stock Price Prediction in Indonesia Companies," *Sci. J. Informatics*, vol. 11, no. 1, pp.





- 61–68, 2024, doi: 10.15294/sji.v11i1.47935.
- [9] F. Rumaisa, Y. Puspitarani, A. Rosita, A. Zakiah, and S. Violina, “Penerapan Natural Language Processing (NLP) di bidang pendidikan,” *J. Inov. Masy.*, vol. 1, no. 3, pp. 232–235, 2021, doi: 10.33197/jim.vol1.iss3.2021.799.
- [10] F. I. Muqsih, E. Supriyati, and T. Listyorini, “Klasifikasi Pengucapan Huruf Hijaiyah Berbasis Android Menggunakan CNN dengan Fitur Mel-Spectrogram,” *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 67–78, 2025, doi: 10.30591/jpit.v10i1.8145.
- [11] S. E. Sitepu, “Low-Resource Single-Domain Machine Translation untuk Bahasa Karo-Indonesia,” *Indones. J. Multidisciplinary Expert. J. Multidisiplin Ilmu*, vol. 1, no. 4, p. 2023, 2023.
- [12] D. F. Wiguna, Z. F. Fachrina, and Al. H. Al Hadi, “PERKEMBANGAN SISTEM TERJEMAH OTOMATIS BERBASIS KECERDASAN BUATAN: STUDI ANALISIS TERHADAP GOOGLE TRANSLATE DALAM MENERJEMAHKAN TEKS ARAB-INDONESIA,” *Al-Furqan J. Agama, Sos. dan Budaya*, vol. 4, no. 3, pp. 892–901, 2025, [Online]. Available: <https://publisherqu.com/index.php/Al-Furqan>
- [13] M. R. Wijaya, “Inovasi Model Intrusion Detection System (IDS) menggunakan Double Layer Gated Recurrent Unit (GRU) dengan Fitur Berbasis Fusion,” *J. Ilm. Educic Pendidik. dan Inform.*, vol. 12, no. 1, pp. 10–21, 2025, doi: <https://doi.org/10.21107/educic.v12i1.28822> Diterima:
- [14] D. I. Af'idah and A. Susanto, “Hyperparameter Tuning Seq2Seq Gated Recurrent Unit Untuk Penerjemahan Bahasa Daerah Ke Nasional,” *J. Inform. Teknol. dan Sains*, vol. 6, no. 4, pp. 1238–1248, 2025, doi: 10.51401/jinteks.v6i4.5645.
- [15] A. Pertiwi, H. Sujaini, and N. Safriadi, “Pengaruh Penggunaan Attention pada Kualitas Mesin Penerjemah Berbasis Neural Bahasa Indonesia ke Bahasa Madura Bangkalan,” *J. Linguist. Komputasional*, vol. 7, no. 1, pp. 11–20, 2024, doi: 10.26418/jlk.v7i1.197.
- [16] F. Razsiah, A. Josi, and S. Mubaro, “Aplikasi Penerjemah Bahasa Bangka Ke Bahasa Indonesia Menggunakan Neural Machine Translation Berbasis Website,” *J. Inov. Teknol. Terap.*, vol. 1, no. 1, pp. 68–76, 2023, doi: 10.33504/jitt.v1i1.67.
- [17] W. Gunawan, H. Sujaini, and T. Tursina, “Analisis Perbandingan Nilai Akurasi Mekanisme Attention Bahdanau dan Luong pada Neural Machine Translation Bahasa Indonesia ke Bahasa Melayu Ketapang dengan Arsitektur Recurrent Neural Network,” *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. 488, 2021, doi: 10.26418/jp.v7i3.50287.
- [18] Calvin, Piter Antonius, and Saut Dohot Siregar, “Google Translate Mengubah Bahasa Batak Menjadi Bahasa Indonesia,” *Repeater Publ. Tek. Inform. dan Jar.*, vol. 3, no. 2, pp. 08–16, 2025, doi: 10.62951/repeater.v3i2.399.
- [19] A. Wijanarko and A. N. Al Haura, “Model Recurrent Neural Network-Gated Recurrent Unit untuk Membangun Mesin Penerjemah Bahasa Indonesia-Banyumasan,” *J. Eksplora Inform.*, vol. 13, no. 2, pp. 218–226, 2024, doi: 10.30864/eksplora.v13i2.977.
- [20] D. M. Nugraha and D. Ariatmanto, “Kata Kunci : Prediksi ; Hibrid ; GRU-LSTM ; Bitcoin ; Data Analisis,” vol. 11, no. April, 2025.
- [21] D. Nuraisyah, Sumarni, N. A. Riana, and A. Syahid, “Menyelidiki Pola Kolokasi dalam Terjemahan Bahasa Inggris dan Bahasa Indonesia dari The Litle Prince : Studi Korpus,” *J. Pendidik. Bhs.*, vol. 15, pp. 97–102, 2025.
- [22] R. M. Adani, P. P. Adikara, and N. Santoso, “Analisis Klaster Terjemahan Ayat Al- Qur ’ an Berbahasa Indonesia Menggunakan K-Means dan Word Embedding,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 9, pp. 1–8, 2025.
- [23] N. Method, M. A. Pratama, and T. Y. Hadiwandura, “Implementasi Deep Learning Untuk Identifikasi Jenis Biji Kopi Menggunakan Metode Convolutional Neural Network,” *CSRID J.*, vol. 17, no. 3, pp. 387–398, 2025.
- [24] M. Jayanti, I. Purnama, S. Teknologi, P. Studi, S. Informasi, and U. Labuhanbatu, “Rancang Bangun Aplikasi Data Guru Berbasis React Native,” *Gemilang Inform.*, vol. 3, no. 1, pp. 24–31, 2025.
- [25] I. D. Mienye, T. G. Swart, and G. Obaido, “Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications,” *Inf.*, vol. 15, no. 9, pp. 1–34, 2024, doi: 10.3390/info15090517.
- [26] E. S. Bere and Y. R. Kaesmetan, “Perbandingan Kinerja Convolutional Neural Network , Recurrent Neural Network Dan Long Short Term MEMORY DALAM ANALISIS SENTIMEN,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 4, 2025.
- [27] M. A. Syifa, D. Retno, and S. Saputro, “Stance Detection Dengan Algoritme Gated Recurrent Unit





- (GRU),” *Pros. Semin. Nas. Mat. Stat. dan Apl.*, pp. 267–275, 2023.
- [28] M. Hafiz, R. Pratama, M. Akrom, A. P. Santosa, and M. Reesa, “Klasifikasi Otomatis Korosi Menggunakan Convolutional Neural Network dan Transfer Learning dengan Model MobileNetV2,” *J. Algoritm.*, vol. 22, pp. 138–148, 2025, doi: 10.33364/algoritma/v.22-1.2182.
- [29] F. Shahira, B. S. Negara, F. Yanto, and S. Sanjaya, “Optimasi Hyperparameter Deep Learning untuk Deteksi X-,” *JIEET (Journal Inf. Eng. Educ. Technol.)*, vol. 09, pp. 56–63, 2025.
- [30] M. Huzaiifa *et al.*, “Model IndoBERT Dalam Pengenalan Kebudayaan Jawa Berbasis Question Answering Abstrak,” in *SNIV: SEMINAR NASIONAL INOVASI VOKASI eISSN*, 2025, vol. 4, no. 1, pp. 416–423.
- [31] N. Nyoman, E. Perimawati, R. R. Huizen, and D. P. Hostiadi, “Analisa Pengaruh Pre-Processing Data Untuk Model Deteksi Akun Palsu Pada Media Sosial,” in *SPINTER: Prosiding Seminar Hasil Penelitian Informatika dan Komputer*, 2025, vol. 2, no. 1, pp. 1117–1122.
- [32] F. Ashfaq, N. Z. Jhanjhi, and N. A. L. I. Khan, “Enhancing ECG Report Generation With Domain-Specific Tokenization for Improved Medical NLP Accuracy,” *IEEE Access*, vol. 13, no. April, pp. 85493–85506, 2025, doi: 10.1109/ACCESS.2025.3567566.
- [33] Adriyendi, *ARTIFICIAL INTELLIGENCE dengan model diskriminatif dan model generatif*. Bandung: WINDINA MEDIA UTAMA, 2025.
- [34] E. Supriyati, “Recognition System of Indonesia Sign Language based on Sensor and Artificial Neural Network,” *Makara J. Technol.*, vol. 17, no. 1, 2013, doi: 10.7454/mst.v17i1.1924.
- [35] B. Imran and A. Hidayat, “Implementasi Machine Learning untuk Mendeteksi Penyakit Katarak menggunakan Kombinasi Ekstraksi Fitur dan Neural Network Berdasarkan Citra,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 7, no. 2, pp. 232–251, 2025, doi: <https://doi.org/10.35746/jtim.v7i2.621>.