5. PENGUJIAN

Pada Bab ini akan dibahas mengenai pengujian terhadap model yang telah ditentukan pada saat proses *training* dan *testing*. Tahap ini meliuti pengujian terhadap *data training* dari *dataset* yang ada dan juga terhadap model yang telah dibuat. Pengujian ini dilakukan dengan Google Colab memalui laptop ROG NVIDIA GeForce GTX 1660Ti dengan intel Core i7-10750 dengan RAM 8.

5.1. Perangkat Lunak yang Digunakan

Pembuatan Interface pada skripsi ini menggunakan bantuan *gradio* dan pengujian data training dengan menggunakan *Python*. Yang mana gradio dibuat pada Visual Studio Code dan training data dilakukan di Google Colab.

5.2. Tujuan Pengujian

Pengujian ini dilakukan untuk mendapatkan hasil yang terbaik dari parameter dan cara penerjemahan yang berbeda agar dapa menghadilkan kualitas penerjemahan yang terbaik. Pengujian juga dilakukan untuk memastikan metode dapat digunakan dalam penerjemahan. Kemudian, model dengan parameter terbaik akan digunakan dakam pembuatan *website*.

5.3. Pengujian Model

Pengujian ini akan dilaklukan pada Model Bi-LSTM. Pengujian ini dilakukan pada algoritma Bi-LSTM dengan penerjemahan bersama IndoBERT dan tanpa IndoBERT. Tedapat juga beberapa pengujian lain yaitu, menguji model dengan jumlah data yang berbeda. Pengujian akan dilakukan dengan membandingkan hasil akurasi dan hasil BLEU Score yang dihasilkan. Pengujuan dilakukan dengan pembagian rasio data sebanya 80% untuk *training* dan 20% lainnya untuk *testing*.

5.3.1. Pengujian Jumlah Data

Pada pengujian ini diberikan jumlah data yang berbeda kepada 1 model yang sama dimana nantinya akan Model Bi-LSTM akan memiliki layer yang sama yaitu *Embedding, Bidirectional(128), Bidirectional(64), GlobalAveragePooling1D(), Dropout(0.5), Dense(128, activation='relu'), Dense(vocab_size+1, activation='softmax')]).* Model akan dilatih dengan jumlah epoch yang sama yaitu 10. Kalimat yang diuji untuk menghitung BLEU Score adalah ‘Titiang ngadol sanganan’ artinya ’saya menjual makanan.

Tabel 5.1.

Tabel Hasil *Training Data* Berdasarkan Jumlahnya

	Jumlah Data	**Akurasi**	**BLEU Score**
Bi-LSTM tanpa IndoBERT	20000	0.7008	0.11362193664674995
Bi-LSTM dengan IndoBERT	20000	0.7008	0
Bi-LSTM tanpa IndoBERT	25000	0.4476	0.11362193664674995
Bi-LSTM dengan IndoBERT	25000	0.4476	0
Bi-LSTM tanpa IndoBERT	15000	0.4484	0.11362193664674995

Dengan melihat hasil *training* pada Tabel 5.1, Bi-*LSTM* dengan menggunakan *IndoBERT* tidak memiliki hasil yang memuaskan sehingga pada pengujian selanjutnya diputuskan untuk tidak menggunakan IndoBERT lagi.

5.3.2. Pengujian Kalimat Terhadap Model

Pada bagian ini dilakukan pengujian terhadap model Bi-LSTM dan Bi-LSTM dengan transformer untuk melihat hasil dari penerjemahan kalimat terhadap model yang telah dibuat. Dengan menggunakna jumlah data yang sama yaitu 60.000 data dan model Bi-LSTM yang sama yaitu, *Embeding, Bidirectional(520), Bidirectional(256), GlobalAveragePooling1D, Dropout(0.5), Dense(512), dan Dense(vocab_size+1, activation='softmax')*, dengan akurasi sebesar 0,7558. IndoBERT nantinya akan digunakan dengan pemanggilan T5 dan *tokenizer IndoBERT*. Dengan jumlah epoch sebanyak 10.

Tabel 5.2.

Tabel Hasil Penerjemahan Kalimat Terhadap Model

Kalimat dalam Bahasa Bali	**Hasil Kalimat Dalam Bahasa Indonesia**	**Hasil Terjemahan Kalimat dengan Bi-LSTM**	**Hasil Terjemahan Kalimat Bi-LSTM dengan IndoBERT**

ratu seneng ngrayunang adeng	ratu seneng makan telor	tunggu senang makan bahan yang terbuat dari kayu bambu tempurung kelapa yang dibakar setelah gosong disiram air untuk didinginkan agar menjadi arang yang siap digunakan sebagai bahan bakar	tunggu tunggu tunggu tunggu tunggu tunggu tunggu
dayu putu ngadol sanganan	Dayu putu menjual kue	tunggu tunggu menjual kue	tunggu tunggu tunggu tunggu tunggu tunggu tunggu
wong jerone sering pisan adua	Pelayan itu sering sekali berbohong	diacunginya ular tunggu tunggu sekali berbohong	tunggu tunggu tunggu tunggu tunggu tunggu tunggu

5.3.3. Pengujian Model Bi-LSTM

Pada bagian ini, akan dilakukan uji coba kepada model Bi-LSTM berdasarkan *layer* yang dimiliki. Data yang digunakan akan memiliki nilai yang sama yaitu 60.000 data. Pengujian ini juga hanya akan menggunakan nilai epoch yang sama yaitu 10. Kemudian pada Penelitian ini akan mengubah layer *bidirectional* dan *dense* saja. Hal ini dilakukan untuk melihat banyak *layer* yang bagus untuk digunakan pada jumlah data yang besar. Nantinya model akan memiliki beberapa layer seperti, *Embedding, Bidirectional, GlobalAveragePooling1D, Dropout,* dan Dense.

Tabel 5.3

Tabel Hasil Pengujian Model Bi-LSTM dengan Julah Layer Yang Berbeda

Layer Bidirectional_1	*Layer Bidirectional_2*	*Layer Dense*	*Akurasi*	*BLEU Score*
512	*256*	*512*	*0.79*	*0.11362193664674995*
512	*256*	*256*	*0.7709*	*0.030516610155986567*

512	*256*	*128*	*0.5978*	*0.018796702320784967*
256	*128*	*256*	*0.7241*	*0.16821895003341453*
256	*256*	*256*	*0.6918*	*0.16821895003341453*

5.3.4. Pengujian dengan *Dataset* Yang Berbeda

Pada Pengujian ini, dilakukan pengujian terhadap 2 *dataset* yang berbeda untuk melihat bagaimana pengaruh bentuk dataset yang berbeda terhadap *model*. Pengujian ini dilakukan dengan jumlah *dataset* yang sama yaitu 60.000. juga menggunakan model *bi-LSTM* yang sama yang terdiri dari beberapa layer yaitu, *Embeding, Bidirectional(520), Bidirectional(256), GlobalAveragePooling1D, Dropout(0.5), Dense(512), dan Dense(vocab_size+1, activation='softmax')*. Pengujian setidap *dataset* akan dilakukan dengan jumlah *epoch* yang sama yaitu 10.

Tabel 5.4.

Hasil Pengujian Terhadap *Dataset*

Dataset	**Akurasi**	***BLEU Score***
Dataset Kalimat	0.81	0.11362193664674995
Dataset TERM	0.9494	0

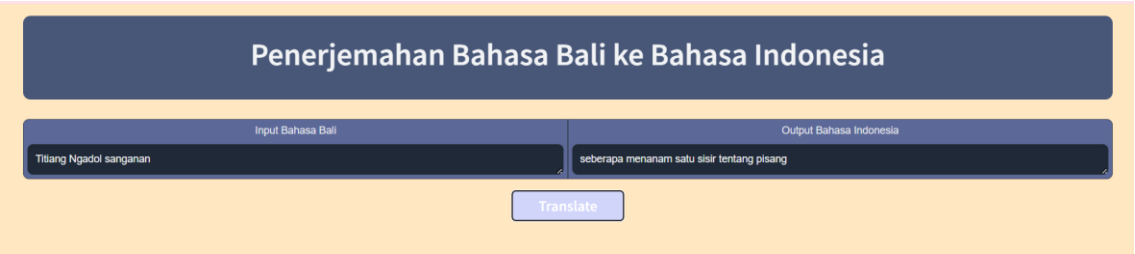
Dapat dilihat dari hasil pengujian pada tabel 5.3. *BLEU Score* dari *dataset* kalimat sangatlah kecil bahkan memiliki nilai 0 meskipun memiliki akurasi yang lebih tinggi dibandingkan dengan *dataset TERM*. SEhingga pada penelitian ini digunakan *dataset TERM* pada setiap pengujian.

5.4. Pengujian Jawaban

Pengujian ini dilakukan terhadap *interface* untuk melihat bagaimana penerjemahan berjalan di *interface*. Pengujian akan dilakukan dengan menggunakan 5 kalimat Bahasa Bali dan akan dilihat output yang dikeluarkan. Berikut adalah beberapa hasil dari pengujian kalimat. Pengujian Jawaban dilakukan dengan menggunakan model yang telah dilatih sebelumnya.

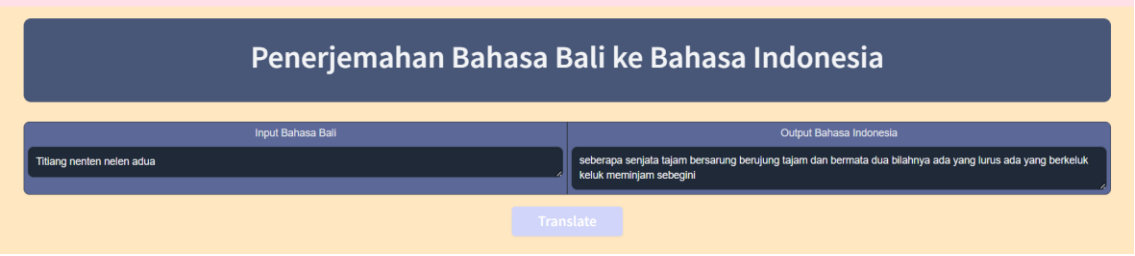
Gambar 5.1. Gambar Pengujian Jawaban Terhadap Kalimat

Pada Gambar 5.1 dilakukan percobaan terhadap kalimat dalam Bahasa bali “Ratu seneng ngrayunang adeng” dan mengeluarkan *output* “meminjam sungguhsungguh upacara pembakaran mayat saya”. *Output* tersebut tentu saja bukan jawaban yang benar. Arti yang benar dalam bahasa Indonesia untuk kalimat tersebut adalah “Ratu suka makan telur”.



Gambar 5.2 Gambar Pengujian Terhadap kalimat

Pada Gambar 5.2 dilakukan percobaan terhadap kalimat dalam Bahasa bali “Titiang Ngadol sanganan” dan mengeluarkan *output* “seberapa menanam satu sisir tentang pisang”. *Output* tersebut tentu saja bukan jawaban yang benar. Arti yang benar dalam bahasa Indonesia untuk kalimat tersebut adalah “Saya menjual makanan”.



Gambar 5.3 Gambar Pengujian Terhadap Kalimat

Pada Gambar 5.3 dilakukan percobaan terhadap kalimat dalam Bahasa bali “Titiang nenten nelen adua” dan mengeluarkan *output* “seberapa senjata tajam bersarung berujung tajam

dan bermata dua bilahnya ada yang lurus ada yang berkeluk keluk meminjam sebegini". *Output* tersebut tentu saja bukan jawaban yang benar. Arti yang benar dalam bahasa Indonesia untuk kalimat tersebut adalah "Saya tidak pernah berbohong".

5.5. Diskusi

Menurut hasil dari pengujian diatas Model yang paling baik untuk digunakan susunannya adalah sebagai berikut:

- *Embedding()*
- *Bidirectional(512)*
- *Bidirectional(256)*
- *GlobalAveragePooling1D(),*
- *Dropout(0.5),*
- *Dense(256, activation='relu'),*
- *Dense(vocab_size+1, activation='softmax')*
- Hasil dari Model yang digunakan memiliki akurasi yang besar daripada yang lain yaitu 79%.
- Meski memiliki hasil akurasi yang tinggi hasil dari penerjemahan masih belum benar dan tidak sesuai. Hal ini bisa saja dikarenakan beberapa hal yaitu salah satunya adalah terlalu banyaknya *noise* didalam atau ataupun adanya data i*mbalance.*