

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia memiliki beragam suku, ras dan budaya dengan penggunaan bahasa daerah yang berbeda-beda. Salah satu bahasa daerah yang ada di Indonesia adalah Bahasa Madura. Bahasa Madura merupakan bahasa utama yang biasa digunakan oleh Suku Madura yang berada di Pulau Madura dan pulau-pulau kecil di sekitarnya. Saat ini, Bahasa Madura tidak hanya digunakan oleh masyarakat Pulau Madura saja. Namun, telah berkembang mencakup banyak wilayah. Beberapa wilayah di luar Pulau Madura yang menggunakan Bahasa Madura sebagai bahasa utamanya adalah Banyuwangi, Situbondo, Bondowoso, hingga Kabupaten Jember [1].

Pada umumnya masyarakat Suku Madura memiliki kemampuan untuk menguasai dua bahasa atau yang dapat disebut dwibahasa [1]. Untuk bahasa utama yang banyak digunakan adalah Bahasa Madura itu sendiri, karena bahasa tersebut merupakan bahasa ibu. Kemudian untuk bahasa kedua merupakan Bahasa Indonesia. Penggunaan Bahasa Indonesia sebagai Bahasa Nasional yang digunakan secara umum oleh masyarakat Pulau Madura menyebabkan penggunaan Bahasa Madura tergeserkan dengan Bahasa Indonesia. Dan keberadaan masyarakat pendatang yang sama sekali tidak memahami Bahasa Madura mengakibatkan masalah komunikasi antara masyarakat yang hanya dapat menggunakan Bahasa Madura dan masyarakat yang hanya bisa menggunakan Bahasa Indonesia [2].

Hal ini menyebabkan penggunaan Bahasa Madura tergeser oleh Bahasa Indonesia, terutama oleh masyarakat pendatang yang tidak memahami Bahasa Madura sama sekali [2]. Hal ini mengakibatkan masalah komunikasi antara masyarakat Madura yang hanya menggunakan Bahasa Madura dan masyarakat yang hanya bisa menggunakan Bahasa Indonesia.

Untuk mengatasi masalah tersebut, dibutuhkan jembatan untuk menghubungkan antara Bahasa Indonesia dengan Bahasa Madura. Salah satu solusinya adalah penggunaan alat penerjemah yang memungkinkan untuk menerjemahkan Bahasa Indonesia dengan Bahasa Madura. Untuk pembuatan alat penerjemah tersebut dapat menggunakan suatu mesin yang disebut mesin

penerjemah. Mesin penerjemah adalah mesin yang dapat melakukan proses penerjemahan dari satu bahasa ke bahasa lainnya secara otomatis. Sehingga dapat membantu pengguna untuk melakukan komunikasi dengan bahasa yang berbeda [3].

Mesin penerjemah termasuk salah satu pendekatan dari kecerdasan buatan *Natural Language Processing* (NLP). Tujuan utama dari NLP adalah membuat mesin yang dapat memproses bahasa yang digunakan manusia menjadi dapat dipahami atau diterima oleh komputer, sehingga komputer dapat memberikan respon yang sesuai [4]. Dari beberapa penelitian yang telah dilakukan dalam pemanfaatan NLP untuk mesin penerjemah. Terdapat salah satu metode yang dapat digunakan untuk merancang mesin penerjemah, yaitu *Neural Machine Translation* (NMT). NMT merupakan metode dalam pembuatan mesin penerjemah yang memanfaatkan model *Recurrent Neural Network* (RNN) dengan arsitektur *encoder* dan *decoder* [5]. *Encoder* merupakan tahapan pada RNN yang mengkodekan input menjadi representasi vektor panjang tetap. Setelah *encoder* menghasilkan representasi vektor dari input, *decoder* bertugas menggunakannya untuk menghasilkan urutan keluaran yang diinginkan [5].

Dalam pengembangannya, RNN memiliki dua variasi yang dibuat untuk mempertahankan masalah informasi jangka panjang yaitu *Long Short-Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU) [6]. Dalam mengatasi masalah yang dimiliki oleh RNN standar, LSTM menggunakan pendekatan yang lebih rumit dengan menambahkan unit memori dan tiga jenis *gate*, yaitu *input gate*, *forget gate* dan *output gate* [6]. Dalam hal tersebut, LSTM memiliki kelebihan dalam kemampuan pengolahan dan penyimpanan informasi, namun juga memiliki sifat yang intensif dalam hal komputasi dan penggunaan memori. Pada sisi lain, pengembangan GRU memiliki dampak positif dalam hal efektivitas. Berbeda dengan LSTM, GRU memiliki model yang lebih sederhana dengan hanya memiliki dua *gate*, yaitu *update gate* dan *reset gate* [6]. Meskipun hanya memiliki dua *gate*, GRU dinilai mampu mengatasi masalah hilangnya informasi jarak jauh dan memiliki waktu komputasi yang lebih singkat.

Pada penelitian sebelumnya mengenai mesin penerjemah untuk bahasa indonesia - bahasa madura penelitian pertama kali dilakukan pada tahun 2013

menggunakan metode *Free Context Parsing Algorithm* [7], kemudian pada tahun 2014 dilakukan penelitian serupa menggunakan metode *Rule Based* [8] dan untuk penelitian terbaru mengenai penerapan mesin penerjemah bahasa Madura – bahasa Indonesia menggunakan NMT dengan Model *Transformer* [9]. Dikarenakan penerapan mesin penerjemah Bahasa Indonesia – Bahasa Madura menggunakan model GRU belum pernah dilakukan, maka penerapan tersebut belum dapat diketahui skor BLEU terbaik apabila Bahasa Madura-Indonesia digunakan sebagai obyek mesin penerjemah menggunakan model GRU. Namun terdapat penelitian yang berkaitan dengan hal tersebut, terdapat penelitian mengenai beberapa model untuk membuat mesin penerjemah [10]. Diantara model yang diteliti terdapat model LSTM dan GRU. Dari hasil penelitian tersebut dengan *epoch* sebesar 50, GRU memiliki waktu komputasi 107 *second* yang lebih cepat dibandingkan LSTM dengan waktu komputasi 132 *second*. Meskipun nilai akurasi yang dimiliki LSTM sedikit lebih tinggi dari GRU yaitu 0.940 untuk LSTM dan 0.937 untuk GRU. Hal tersebut dapat membuktikan bahwa GRU merupakan salah satu metode yang efektif digunakan untuk mesin penerjemah.

Maka dari itu dilakukan penelitian dengan judul “Mesin Penerjemah Bahasa Indonesia – Bahasa Madura Menggunakan Model *Gated Recurrent Unit*”. Penelitian ini bertujuan untuk merancang mesin penerjemah untuk menerjemahkan Bahasa Indonesia – Bahasa Madura dan meningkatkan minat terhadap penggunaan Bahasa Madura.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas maka dapat dipaparkan rumusan masalah sebagai berikut:

1.2.1 Permasalahan

Pengembangan NMT untuk mesin penerjemah Bahasa Indonesia - Bahasa Madura.

1.2.2 Usulan Solusi

Solusi yang ditawarkan dalam penelitian ini adalah mengembangkan metode NMT untuk mesin penerjemah Bahasa Indonesia – Bahasa Madura adalah menggunakan *dataset* yang berisi kalimat Bahasa Indonesia – Bahasa Madura. Untuk preprocessing menggunakan empat tahap, yaitu *case folding*, *punctuation*

removal, *tokenization*, dan *add padding*. Untuk metodenya menggunakan model *Gated Recurrent Unit* (GRU). Dan *Bilingual Evaluation Understudy* (BLEU) digunakan untuk melakukan evaluasi terhadap model.

1.2.3 Pertanyaan Peneliti

Berapa skor BLEU terbaik yang diperoleh model GRU dalam melakukan penerjemahan Bahasa Indonesia – Bahasa Madura?

1.3 Tujuan dan Manfaat

Adapun tujuan dan manfaat dari penelitian ini adalah

1.3.1 Tujuan Penelitian

Penelitian ini bertujuan untuk mengetahui skor BLEU terbaik yang diperoleh Mesin Penerjemah Bahasa Indonesia – Bahasa Madura menggunakan model GRU

1.3.2 Manfaat Penelitian

Dengan penerapan mesin penerjemah Bahasa Indonesia – Bahasa Madura menggunakan model GRU diharapkan menjadi alat yang dapat digunakan untuk membantu dalam belajar Bahasa Madura

1.4 Batasan Masalah

Batasan masalah diperlukan guna penelitian tetap terarah dan fokus pada hal yang diteliti, batasan tersebut diantaranya:

1. Dataset yang digunakan berbentuk kalimat paralel Bahasa Indonesia – Bahasa Madura
2. Menggunakan model GRU dengan arsitektur encoder-decoder.
3. Tidak memperhatikan tingkatan atau dialek Bahasa Madura.

1.5 Sistematika Penulisan

Mengenai masalah yang akan diulas oleh penulis, maka penulis menguraikan menjadi beberapa bab yang secara garis besar adalah sebagai berikut:

1. BAB I PENDAHULUAN

Pada bab ini memuat latar belakang tentang penyebaran Bahasa Madura, penggunaan bahasa selain Bahasa Madura di Pulau Madura, pergeseran Bahasa Madura, pemanfaatan mesin penerjemah untuk menjadi jembatan Bahasa Madura dan Bahasa Indonesia, penjelasan singkat mengenai NLP, NMT, dan penjelasan mengenai GRU untuk mesin penerjemah. Kemudian bab ini memuat rumusan masalah yang berisi

permasalahan, usulan solusi, dan pertanyaan penelitian. Lalu terdapat bab ini juga memuat tujuan dan manfaat penelitian. Selanjutnya terdapat batasan masalah. Dan yang terakhir adalah sistematika penulisan.

2. BAB II KAJIAN PUSTAKA

Pada bab ini dijelaskan mengenai teori teori yang mendukung dan membahas mengenai deep learning, NLP, preprocessing, Word2Vec, NMT, Fungsi Aktivasi, RNN, GRU, dan BLEU, Flask, Figma, dan Penelitian Terkait. Diharapkan dari semua teori yang tersebut dapat membantu dalam penelitian.

3. BAB III METODE PENELITIAN

Pada bab ini berisi mengenai tahapan-tahapan penelitian mengenai pembuatan mesin penerjemah Bahasa Madura – Bahasa Indonesia menggunakan model GRU.

4. BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi mengenai lingkungan uji coba yang digunakan, Kode program yang telah dibuat, dan juga hasil dari uji coba seluruh skenario.

5. BAB V KESIMPULAN DAN SARAN

Pada bab ini berisi mengenai kesimpulan semua hasil dari penelitian yang telah dilakukan. Dan berisi tentang saran untuk penelitian ke depannya.