

## INTISARI

# PERINGKASAN BERITA BERBAHASA INDONESIA SECARA ABSTRAKTIF DENGAN REPRESENTASI INPUT GABUNGAN *WORD EMBEDDING* DAN POS TAG MENGGUNAKAN ENCODER-DECODER GATED RECURRENT UNIT

Oleh

Felix Ogustino

17/412636/PA/17955

Data teks merupakan salah satu data yang mudah ditemukan karena ketersediaannya secara publik dan jumlahnya yang banyak. Salah satu contoh data teks yang tersedia secara publik dan jumlahnya banyak adalah data dalam bentuk artikel berita daring. Ketersediaan artikel berita yang banyak membuat pembaca bisa membaca berita dengan berbagai macam kategori dan sumber. Akan tetapi, artikel berita yang diterbitkan sangat banyak setiap harinya, sehingga dibutuhkan waktu yang lama jika ingin mengekstrak informasi dari banyak artikel berita tersebut. Peringkasan teks otomatis merupakan salah satu alternatif cara yang dapat membantu dalam menghadapi permasalahan tersebut.

Penelitian ini membuat suatu peringkasan berita berbahasa Indonesia secara abstraktif dengan representasi input gabungan *word embedding* dan POS Tag menggunakan Encoder-Decoder Gated Recurrent Unit pada *dataset* artikel berbahasa Indonesia. Untuk mengevaluasi performa dari metode tersebut, akan digunakan metode ROUGE untuk membandingkan kualitas ringkasan yang dibuat secara otomatis dengan ringkasan yang dibuat manual oleh manusia.

Berdasarkan pengujian yang telah dilakukan, model memperoleh skor f-measure ROUGE-1 sebesar 9,63%, ROUGE-2 sebesar 0,10%, dan ROUGE-L sebesar 8,46% pada skenario pengujian dengan representasi input hanya *word embedding* saja. Untuk skenario pengujian dengan tambahan representasi input POS Tag, model memperoleh skor f-measure ROUGE-1 sebesar 10,72%, ROUGE-2 sebesar 0,16%, dan ROUGE-L sebesar 9,47%.

Kata Kunci: Peringkasan Teks Otomatis, Abstraktif, GRU

## ABSTRACT

### ABSTRACTIVE SUMMARIZATION OF INDONESIAN NEWS USING COMBINED WORD EMBEDDING AND POS TAG INPUT REPRESENTATION WITH ENCODER-DECODER GATED RECURRENT UNIT

By

Felix Ogustino

17/412636/PA/17955

Text is one of the most common types of data that is easy to find because of its public availability and high volume. Online news articles are one example of text data that is widely accessible and available in large quantities. The availability of news articles allows readers to read news from various categories and sources. However, a large number of news articles are published every day, making the process of extracting information from numerous news articles time-consuming. The use of automatic text summarization is an alternative approach that can be used to address these issues.

This study uses a combined word embedding and POS Tag input representation with Encoder-Decoder Gated Recurrent Unit to construct an Abstractive Summarization of Indonesian News to process a dataset that consists of Indonesian news articles. The ROUGE approach will be used to compare the quality of the summary produced automatically with the summary produced manually in order to assess the effectiveness of the method.

Based on the result of the evaluation, the model get f-measure score 9.63% for ROUGE-1, 0.10% for ROUGE-2, and 8.46% for ROUGE-L in the testing scenario using only word embeddings as input representation. When POS Tag input representation was added in the testing scenario, the model get f-measure score 10.72% for ROUGE-1, 0.16% for ROUGE-2, and 9.47% for ROUGE-L.

Keywords: : Automatic Text Summarization, Abstractive, GRU