

## INTISARI

### **ANALISIS SENTIMEN DENGAN *RECURRENT NEURAL NETWORK* DAN *GATED RECURRENT UNIT* PADA DATA IMBALANCED**

Oleh

Rana Aribah

17/409527/PA/17834

Media sosial merupakan wadah yang biasa digunakan untuk mengungkapkan suatu opini. Salah satu media sosial yang populer adalah *twitter*. Sosial media dapat menjadi wadah perusahaan berbagi informasi dan mendapat *feedback* dari konsumen dengan cara yang mudah dan cepat. Kebutuhan masyarakat akan layanan internet terus bertambah. First Media merupakan salah satu perusahaan yang menyediakan jasa tersebut. Bagi perusahaan, opini konsumen penting dianalisis untuk evaluasi dan pengembangan produk. Diperlukan metode untuk mengklasifikasikan tanggapan ke dalam sentimen secara otomatis. Pada penelitian ini menggunakan data *tweet* tentang First Media, dimana *tweet* merupakan data sekuensial, yaitu data yang mempunyai urutan. Metode yang dapat memproses data sekuensial adalah *Recurrent Neural Network* dan *Gated Recurrent Unit*. Pada penelitian ini dilakukan teknik *random oversampling* untuk menangani data tidak seimbang. Berdasarkan eksperimen dimensi *word embedding*, jumlah neuron, fungsi aktivasi, dan penggunaan *dropout* pada data latih yang seimbang dan tidak seimbang menggunakan kedua metode tersebut, diperoleh hasil bahwa pembentukan model pada metode *Gated Recurrent Unit* dengan pelatihan menggunakan data tidak seimbang mempunyai performa yang paling baik dengan rata-rata akurasi sebesar 93,27%, presisi sebesar 76,02%, sensitivitas sebesar 61,56%, spesifisitas sebesar 97,45%, dan f-1 score sebesar 67,90%.

Kata kunci : analisis sentimen, *recurrent neural network*, *gated recurrent unit*, *random oversampling*

## ABSTRACT

### ***SENTIMENT ANALYSIS USING RECURRENT NEURAL NETWORK AND GATED RECURRENT UNIT FOR IMBALANCED DATA***

By

Rana Aribah

17/409527/PA/17834

Social media is a platform which is regularly used to express opinion(s). One of the well-known social media platforms is Twitter. The implication of social media are vary, one of them are for companies to share information and provide feedback from customers in an effortless and fast way. Consequently, this demand will increase the public's need for internet service. First Media is one of the leading companies that provides internet service in Indonesia. In order to evaluate their service and to develop their company, it is necessary to analyze customers opinion. A new method is needed to automatically classify responses into sentiments. This study is conducted using tweet about First Media, in which involving sequential data where the data have order. Methods that are able to process sequential data are Recurrent Neural Networks and Gated Recurrent Units. In our study, random oversampling is applied to handle imbalanced data. Based on experiments word embedding dimensions, number of neurons, activation function, and dropout usage on balanced and imbalanced data training using both methods, it is found that the modeling of the Gated Recurrent Unit method with training using imbalanced data have the best performance with mean of the accuracy is 93.27%, precision 76.02%, sensitivity 61.56%, specificity 97.45%, and f-1 score 67.90%.

Keywords : sentiment analysis, recurrent neural network, gated recurrent unit, random oversampling