

PEMODELAN TOPIK TERKAIT BANJIR PADA TWITTER DENGAN MENGUNAKAN LATENT DIRICHLET ALLOCATION

Muhammad Sutan Irwansyah, Muhammad Habibi, Aris Wahyu Murdiyanto

INTISARI

Latar Belakang : Pada latar belakang ini membahas mengenai topik dari *tweet* yang mengenai Banjir pada Twitter dengan menggunakan kata kunci “Banjir”. Data *tweet* diambil dari rentang tanggal 1 Juni 2021 sampai dengan tanggal 2 Juni 2021 dengan jumlah data *tweet* yang diperoleh yaitu sebanyak 2000 *tweet*. Banyaknya *tweet* yang terkait dengan dengan banjir yang belum dilakukan analisis sehingga belum diketahui topik-topik yang terkandung di dalamnya.

Tujuan Penelitian : Membuat pemodelan topik terkait banjir di wilayah Indonesia pada media sosial Twitter dengan metode LDA.

Metode Penelitian : Penelitian ini menggunakan metode eksperimen dengan beberapa variabel untuk menguji hipotesis. Kemudian data diolah dengan tahap yaitu *web data extraction*, *preprocessing*, *feature extraction*, *topic modelling* menggunakan algoritma *latent dirichlet allocation*, visualisasi, dan analisis.

Hasil Penelitian : Hasil tahap *topic coherence* dilakukan pencarian topik yang paling optimal dari 20 topik yang sudah di tentukan diawal. Hasil *topic coherence* untuk 20 topik disimpulkan bahwa untuk topik 10 memiliki jumlah nilai topik sebesar 0,41 dan memiliki hasil *topic modelling* yang ideal dan sesuai dengan ketentuan.

Kesimpulan : Berdasarkan dari hasil pembahasan *topic coherence*, dapat disimpulkan bahwa jumlah topik yang paling ideal adalah topik 10 dikarenakan memiliki nilai yang paling tinggi dibanding topik lainnya. Saran disini adalah dapat menampilkan atau mendapatkan informasi banjir di Indonesia secara realtime dan akurat.

Kata Kunci: Banjir, LDA, *Topic Modelling*, *Text Mining*, *Latent Dirichlet Allocation*

TOPIC MODELLING FLOOD RELATED TOPICS ON TWITTER USING LATENT DIRICHLET ALLOCATION

Muhammad Sutan Irwansyah, Muhammad Habibi, Aris Wahyu Murdiyanto

ABSTRACT

Background : In this background discusses the topic of tweet about Flooding on Twitter using the keyword "Flood". Tweet data was taken from June 1, 2021 to June 2, 2021 with the number of tweet data obtained, which was 2000 tweets. The number of tweets related to flooding has not been analyzed so that the topics contained in it are not yet known.

Research Objectives : Modeling topics related to floods in Indonesia on Twitter social media with the LDA method.

Research Methods : This study uses experimental methods with several variables to test hypotheses. Then the data is processed with stages, namely web data extraction, preprocessing, feature extraction, topic modeling using latent dirichlet allocation algorithms, visualization, and analysis.

Research Results : The results of the topic coherence stage were carried out a search for the most optimal topic from the 20 topics that had been determined at the beginning. The results of topic coherence for 20 topics concluded that for topic 10 it has a total topic value of 0.41 and has an ideal topic modeling result and is in accordance with the provisions.

Conclusion : Based on the results of the discussion of topic coherence, it can be concluded that the most ideal number of topics is topic 10 because it has the highest value compared to other topics. The advice here is to be able to display or get flood information in Indonesia in real time and accurately.

Keywords: Flood, LDA, Topic Modelling, Text Mining, Latent Dirichlet Allocation