

## ABSTRACT

Indonesia is a developing country that is a member of the United Nations (UN) which has ratified the Sustainable Development Goals (SDGs) valid from 2016 to 2030, it has 17 goals divided into 169 targets that have to be accomplished. The media has a significant role in the growth of the SDGs in Indonesia. Media as a distributor of information on the realization of the SDGs. Today's news has not narrowed down the goals of the SDGs to accomplish. So that trend of news topics is still based on certain categories in general, not yet directed to specific topics based on the news presented.

By finding different topics in a collection of news documents, this research offers topic modeling to find trends in SDGs topics in news documents. Topic modeling uses Latent Dirichlet Allocation (LDA) as an algorithm to extract topics. In the topic modeling output, validation is carried out which will measure the performance of the LDA algorithm and increase the coherence score.

The news document will be carried out in the preprocessing stage using stemming with the weighting of words Term Frequency – Inverse Document Frequency (TF-IDF) and not uses stemming with TF-IDF, then topic modeling with LDA is applied in these two conditions by obtaining coherence score at LDA+TF-IDF is 0.6164, the perplexity of -9.2740025 and LDA non-TF-IDF is 0.5436 perplexity with -7.60408. The test to get the optimal coherence score in this study used LDA-Mallet, where the coherence increased from the previous LDA topic modeling to LDA non-TF-IDF of 0.6119 and LDA+TF-IDF critically increased to 0.7591. The higher the coherence score, the better and easier it is to interpret topic modeling.

**Keywords :** SDGs, Media Online, Topic Modeling, Coherence Score, LDA.

## INTISARI

Indonesia merupakan negara berkembang yang tergabung dalam Perserikatan Bangsa-Bangsa (PBB) telah mengesahkan *Sustainable Development Goals* (SDGs) berlaku sejak 2016 hingga 2030, memiliki 17 *goals* terbagi dalam 169 target yang harus dicapai. Media memiliki peranan penting dalam pengembangan SDGs di Indonesia. Media sebagai penyalur informasi realisasi SDGs. Berita pada hari ini belum mengerucut pada *goals* SDGs yang ingin dicapai. Sehingga tren dari topik berita masih berdasarkan kategori tertentu secara umum, belum mengarahkan ke topik-topik khusus berdasarkan berita yang disajikan.

Dengan menemukan topik-topik yang berbeda dalam suatu kumpulan dokumen berita, penelitian ini menawarkan pemodelan topik untuk menemukan tren topik SDGs dalam dokumen berita. *Topic modeling* menggunakan *Latent Dirichlet Allocation* (LDA) sebagai algoritme untuk mengekstrak topik. Pada hasil luaran *topic modeling* dilakukan validasi yang akan mengukur kinerja algoritme LDA dan meningkatkan nilai koherensi.

Dokumen berita akan dilakukan tahap *preprocessing* dengan menggunakan *stemming* dengan pembobotan kata *Term Frequency – Inverse Document Frequency* (TF-IDF) dan tidak melakukan *stemming* dengan TF-IDF, selanjutnya diterapkan pemodelan topik dengan LDA pada dua kondisi tersebut dengan mendapatkan nilai koherensi pada LDA+TF-IDF 0.6164 *perplexity* -9.2740025 dan LDA non TF-IDF 0.5436 *perplexity* dengan -7.60408. Pengujian untuk mendapatkan nilai koherensi yang optimal pada penelitian ini menggunakan LDA-Mallet, dimana koherensi menjadi meningkat dari pemodelan topik LDA sebelumnya menjadi LDA non TF-IDF sebesar 0.6119 dan LDA+TF-IDF sangat meningkat menjadi 0.7591. Semakin tinggi nilai koherensinya maka semakin baik dan mudah diinterpretasi pemodelan topiknya.

**Kata kunci** – SDGs, Media Online, Pemodelan Topik, *Coherence Score*, LDA.