

INTISARI

PEMODELAN TOPIK PADA DATA REVIEW MENGGUNAKAN *TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY* DAN *LATENT DIRICHLET ALLOCATION*

oleh

Arung Rizky Anenurat
16/394187/PA/17278

Dalam upaya membatasi pertumbuhan Covid-19 dengan mendeteksi alur penyebarannya serta mengetahui setiap kontak erat yang terjadi antara individu, pemerintah melalui Kominfo meluncurkan aplikasi PeduliLindungi. Pada awal diluncurkannya aplikasi PeduliLindungi masih banyak keluhan dari masyarakat mengenai penggunaan aplikasi tersebut. Sehingga aplikasi PeduliLindungi terus diperbaharui untuk memperbaiki kekurangan dan menambah fitur yang masih dianggap kurang. Meski telah mengalami perbaharuan masih banyak keluhan dan saran yang masuk dari masyarakat. Banyaknya kritik dan saran yang disampaikan, tidak mudah bagi pengembang aplikasi PeduliLindungi untuk mengetahui topik apa saja yang dibahas dan perlu menjadi perhatian. Diperlukan metode untuk memodelkan topik apa saja yang dibahas dalam ulasan aplikasi PeduliLindungi pada Google Play Store. Pemodelan topik yang populer digunakan adalah *Latent Dirichlet Allocation* (LDA). Pada skripsi ini penulis menggabungkan metode pemodelan topik LDA dengan algoritma *Term Frequency-Inversed Document Frequency* (TF-IDF). Algoritma TF-IDF merupakan perkalian antara TF dan IDF. Perhitungan estimasi dari distribusi posterior menggunakan algoritma *Variational Expectation Maximization (VEM)*. Hasil pemodelan topik ini berupa proporsi topik pada korpus, probabilitas kata pada setiap topik, dan proporsi topik dalam dokumen/ulasan.

Kata kunci: pemodelan topik, *Latent Dirichlet Allocation*, TF-IDF, *Variational Expectation Maximization*.

ABSTRACT

TOPIC MODELS FOR REVIEW DATA USING TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY AND LATENT DIRICHLET ALLOCATION

by

Arung Rizky Anenurat
16/394187/PA/17278

In an effort to limit the growth of Covid-19 by detecting the flow of its spread and knowing every close contact that occurs between individuals, the government through the Ministry of Communication and Information has launched the PeduliLindung application. At the beginning of the launch of the PeduliLindung application, there were still many complaints from the public regarding the use of the application. So the PeduliLindung application continues to be updated to correct deficiencies and add features that are still considered lacking. Even though it has been updated, there are still many complaints and suggestions from the community. With so many criticisms and suggestions submitted, it is not easy for PeduliLindung application developers to find out what topics are discussed and need attention. A method is needed to model what topics are covered in the PeduliLindungi app review on the Google Play Store. The popular modeling topic used is Latent Dirichlet Allocation (LDA). In this thesis, the author combines the LDA topic modeling method with the Term Frequency-Inversed Document Frequency (TF-IDF) algorithm. The TF-IDF algorithm is a multiplication between TF and IDF. The estimation calculation of the posterior distribution uses the Variational Expectation Maximization (VEM) algorithm. The results of this topic modeling are the proportion of topics in the corpus, the probability of words in each topic, and the proportion of topics in the document/review.

Keywords: topic models, Latent Dirichlet Allocation, TF-IDF, Variational Expectation Maximization.