

INTISARI

Pemodelan Topik dengan Menggunakan *Structural Topic Model* pada Data *Review* Maskapai Garuda Indonesia

Oleh

Cristina Tika Natalia
14/363922/PA/15868

Pada kumpulan data tekstual terdapat berbagai informasi seperti tanggal, nama penulis, gender dan lainnya, yang disebut sebagai metadata. Sering kali metadata tersebut ingin diikutsertakan sebagai kovariat dalam analisis *text mining* salah satunya yaitu pemodelan topik. Namun, analisis pemodelan topik pada umumnya tidak mampu mengakomodasi metadata tersebut. Oleh karena itu, dikembangkan suatu pemodelan topik yang disebut *Structural Topic Model* (STM) yang mampu mengatasi masalah tersebut.

STM merupakan suatu pemodelan topik yang dikembangkan dari model *Latent Dirichlet Allocation* (LDA) dan *Correlated Topic Model* (CTM) yang telah dibahas pada skripsi sebelumnya. Analog dengan kedua model tersebut, estimasi parameter pada model ini menggunakan metode Bayesian dimana nilai estimasi diberikan melalui distribusi posterior dan menggunakan *Variational Expectation Maximization* (VEM) dalam menghitung estimasi distribusi posterior tersebut. Ada 2 komponen dalam model STM yaitu *topical prevalence* yang berfungsi untuk mengetahui seberapa sering topik dalam suatu dokumen didiskusikan dan *topical content* yang mengatur frekuensi suatu kosakata dalam masing-masing topik. Dalam skripsi ini, model STM akan diterapkan untuk data tekstual berupa kumpulan *review* yang diberikan oleh penumpang Garuda Indonesia pada www.airlinequality.com dimana jenis layanan yang disediakan oleh maskapai tersebut menjadi kovariatnya. Kemudian dilakukan perbandingan terhadap ketiga jenis layanan tersebut. Sehingga, selain topik dominan, hasil yang diperoleh dari pemodelan topik tersebut adalah penilaian yang diberikan penumpang Garuda Indonesia berdasarkan jenis layanan yang mereka gunakan relatif sama.

Kata kunci : pemodelan topik, *Structural Topic Model*, *review*, *Variational Expectation Maximization*, *text mining*.

ABSTRACT

Topic Modeling Using Structural Topic Model for Review Data of Garuda Indonesia Airline

by

Cristina Tika Natalia
14/363922/PA/15868

In the textual dataset, there is various information about that data such as date, author's name, gender, and many more. That kind of information is called metadata. Sometimes the metadata wish to be included as a covariate in the text mining analysis, one of that is topic modeling. However, the prior topic modeling analysis is not able to include that metadata. Therefore, a topic modeling called Structural Topic Model (STM) was developed to overcome this problem.

STM is a topic modeling developed from the Latent Dirichlet Allocation and the Correlated Topic Model, which has discussed in the previous thesis. Analogous to those models, estimation of the parameter in this model using the Bayesian method where the estimated value is given by the posterior distribution and Variational Expectation Maximization (VEM) is used to calculate the estimated posterior distribution. Moreover, there is two component in this model, which are topical prevalence, to discover how often the topics are discussed in a document, and topical content, which controls the frequency of the terms in each topic. In this thesis, the STM will be applied to textual data which is a dataset of reviews given by Garuda Indonesia's passengers at www.airlinequality.com where the type of service provided by this airline becomes its covariate. Afterward, the result of each type of service will be compared. So, besides the dominant topic, the result obtained from this topic modeling is the reviews given by Garuda Indonesia's passengers based on the types of service they used are relatively similar.

Key word : *topic modeling, Structural Topic Model, review, Variational Expectation Maximization, text mining.*