

ABSTRAK

Pemodelan Topik Kelompok Jawaban Kuesioner Pertanyaan Terbuka

Desthalia

15/378057/PA/16532

Kuesioner, sebagai salah satu instrumen pengumpulan data, memiliki penggunaan yang sangat luas dalam berbagai bidang ilmu. Pada umumnya, kuesioner memiliki dua format pertanyaan, yaitu pertanyaan terbuka dan pertanyaan tertutup. Dalam menganalisis jawaban kuesioner, dilakukan sebuah proses yang disebut dengan coding. Pada praktiknya, penggunaan kuesioner pertanyaan terbuka terhitung jarang dibandingkan dengan kuesioner pertanyaan tertutup karena perlunya usaha coding yang ekstensif dan memakan waktu untuk menentukan kelompok jawaban, meski jawaban yang dihasilkan kuesioner pertanyaan terbuka dapat menginformasikan peneliti lebih dalam tentang motivasi seorang responden ketika memberikan jawaban tentang sebuah isu.

Pada penelitian ini dilakukan pemodelan topik terhadap jawaban dari kuesioner pertanyaan terbuka menggunakan Latent Dirichlet Allocation (LDA) dan Guided Latent Dirichlet Allocation (GuidedLDA), yang mana masing-masing merupakan pemodelan topik *unsupervised* dan *semi supervised*. Validasi model menggunakan metode uji reliabilitas *Cohen's kappa*. Data yang digunakan dalam penelitian ini adalah data *Community Survey 2016 & 2017* milik pemerintah kota Austin, Texas.

Berdasarkan hasil penelitian, ditemukan alasan mahalanya proses *coding* diakibatkan oleh interpretasi manusia yang bermacam-macam tentang sebuah kelompok jawaban. Baik LDA dan GuidedLDA mampu mengidentifikasi kelompok jawaban yang terdapat dalam data dan masalah yang muncul dari proses identifikasi kelompok jawaban. Adapun GuidedLDA menghasilkan lebih banyak kelompok jawaban dibandingkan dengan LDA. Hasil topik GuidedLDA juga lebih baik dibandingkan dengan LDA karena nilai kappa dari uji reliabilitas milik GuidedLDA lebih tinggi.

Kata kunci: *kuesioner, pertanyaan terbuka, coding, pemodelan topik, Latent Dirichlet Allocation, Guided Latent Dirichlet Allocation*



ABSTRACT

Topic Modeling of Response Categories of Open-ended Questionnaire

Desthalia

15/378057/PA/16532

Questionnaire is an instrument to collect data and used extensively in various fields. There are two well-known format of questionnaire, open-ended and close-ended. To categorize and analyze the responses given to a questionnaire, researchers usually use a procedure called *coding*. However, the use of open-ended questionnaire is comparatively rare compared to close-ended questionnaire because of the need for expensive and time-consuming coding efforts to determine groups of answers, although answers to open-ended questionnaires can provide researchers more nuance about the motivation of a respondent when giving answers to an issue.

In this research, the author applies topic modeling to answers of open questionnaires using Latent Dirichlet Allocation (LDA) and Guided Latent Dirichlet Allocation (GuidedLDA), each is unsupervised and semi supervised topic modeling. The model validation uses the Cohen's kappa reliability test. The data used in this study are Community Survey 2016 & 2017 data done by the city of Austin, Texas.

Based on the results of the study, the root cause of how expensive *coding* efforts for open-ended questionnaire is the varying degree of interpretation of a response category. Both LDA and GuidedLDA are found to be able to identify groups of answers contained in the data and address the challenges that come with identifying response categories. GuidedLDA can produce more response groups compared to LDA. The results of the GuidedLDA topic are also better because the Kappa value of the GuidedLDA is higher than those of LDA.

Keywords: *questionnaire, open-ended, coding, topic modeling, Latent Dirichlet Allocation, Guided Latent Dirichlet Allocation*