

INTISARI

Perbandingan metode Non-negative Matrix factorization dan Latent Dirichlet Allocation pada permasalahan question retrieval untuk studi kasus forum kesehatan

Oleh

Alvin Farkhan Ramadhani

16/394074/PA/17165

Forum tanya jawab kesehatan Alodokter merupakan salah satu media informasi kesehatan yang populer digunakan oleh masyarakat Indonesia. Sistem tanya jawab yang digunakan memiliki kelemahan pada kecepatan dan kesempatan untuk terjawabnya suatu pertanyaan. Untuk menyelesaikan permasalahan tersebut, salah satu solusi yang dapat dilakukan adalah dengan melakukan pencarian dan perbandingan pada pertanyaan yang telah dijawab untuk menjadi rekomendasi bagi pertanyaan baru yang serupa. Besarnya *dataset* dan terbatasnya komputasi membuat perlu dilakukan metode yang tidak membandingkan pertanyaan baru dengan keseluruhan data yang ada pada *dataset*. Pada penelitian ini digunakan metode pemodelan topik untuk memperkecil lingkup data yang akan dilakukan pencarian. Penelitian ini menggunakan model *Latent Dirichlet Allocation* (LDA) dan *Non-negative Matrix Factorization* (NMF) untuk mengelompokkan 148.867 data pertanyaan kesehatan ke dalam 280 topik. Jensen-Shannon *Divergence* digunakan untuk menghitung jarak relevansi antara suatu pertanyaan baru dan 100 data yang memiliki relevansi tertinggi pada topik dominan dari pertanyaan baru tersebut. Lima data yang memiliki jarak relevansi terendah diberikan sebagai rekomendasi pertanyaan serupa. Metode tersebut diimplementasikan pada 50 data tes yang kemudian dilakukan evaluasi oleh dokter dan pengguna non-medis untuk dihitung tingkat akurasi menggunakan *Mean Average Precision* dan *Mean Reciprocal Rank*. Penelitian ini menghasilkan nilai akurasi MAP sebesar 0,488 untuk model LDA dan evaluator dokter, 0,588 untuk model LDA dan evaluator non-medis, 0,556 untuk model NMF dan evaluator dokter, serta 0,584 untuk model NMF dan evaluator non-medis. Sedangkan untuk tipe evaluasi MRR, penelitian ini menghasilkan nilai akurasi sebesar 0,677 untuk model LDA dan evaluator dokter, 0,746 untuk model LDA dan evaluator non-medis, 0,793 untuk model NMF dan evaluator dokter, serta 0,747 untuk model NMF dan evaluator non-medis.

Kata-kata kunci : Pemodelan Topik, *Latent Dirichlet Allocation*, *Non-negative Matrix Factorization*, Rekomendasi Pertanyaan, Pengambilan Pertanyaan, Sistem Tanya Jawab

ABSTRACT

Comparison of Non-negative Matrix Factorization and Latent Dirichlet Allocation on question retrieval problem on health forum case study

By

Alvin Farkhan Ramadhani

16/394074/PA/17165

Alodokter health question and answering forum is one of the popular health information media used in Indonesia. The question and answer system used has a weakness in term of speed and chance for a question to be answered. To solve this problem, one of the possible solution is to search and compare the whole dataset to get the similar question recommendations for new question. The size of the dataset and the limited computational resources makes it required to use method that does not compare new question with entire data contained in the dataset. In this research, topic modelling method is used to minimize the scope of searching and comparing. This research used Latent Dirichlet Allocation (LDA) model and Non-negative Matrix Factorization (NMF) model to cluster 148,867 health question data into 280 topics. Jensen-Shannon Divergence is used to calculate the difference between new question and 100 data that have the highest relevance to the dominant topic of the new question. Five data that have the highest relevance is used as the recommendation. The method is implemented on 50 test data which are then evaluated by doctors and non-medical users to calculate their accuracy using Mean Average Precision and Mean Reciprocal Rank. This study produced an MAP accuracy value of 0.488 for the LDA model and the doctor evaluator, 0.588 for the LDA model and non-medical evaluator, 0.556 for the NMF model and the doctor evaluator, and 0.584 for the NMF model and non-medical evaluator. As for the MRR accuracy, this study produced an accuracy value of 0.677 for the LDA model and the doctor evaluator, 0.746 for the LDA model and non-medical evaluator, 0.793 for the NMF model and the doctor evaluator, and 0.747 for the NMF model and non-medical evaluator.

Keywords : Topic Modelling, Latent Dirichlet Allocation, Non-negative Matrix Factorization, Question Recommendation, Question Retrieval, Question Answering System