

Klasifikasi Domain Spesialisasi Dokter pada Data Teks Forum Tanya Jawab Kesehatan = Classification of Doctor Specialization Domain in Health Question and Answer Forum Text Data

Hendrico Kristiawan, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920535271&lokasi=lokal>

Abstrak

Pertanyaan konsultasi pada sebuah forum daring perlu dijawab oleh dokter spesialis yang tepat agar jawaban yang diberikan akurat dan bermanfaat bagi pengguna yang bertanya. Terkait hal tersebut, penelitian ini membahas tentang pengembangan model yang dapat secara otomatis mengarahkan sebuah pertanyaan konsultasi kesehatan ke dokter dengan spesialisasi yang sesuai. Lebih jauh lagi, model yang dibangun merupakan model klasifikasi multi-label karena sebuah pertanyaan dapat terasosiasi dengan lebih dari satu spesialisasi. Penelitian ini dimulai dengan mengevaluasi keefektifan metode pemetaan berbasis aturan dalam memprediksi data yang dianotasi oleh pakar, dan diperoleh hasil yang menunjukkan tingkat keberhasilan yang cukup. Selanjutnya, dikembangkan sebuah model machine learning yang melakukan klasifikasi domain spesialis dokter. Pelatihan model dilakukan dengan berbagai metode, termasuk supervised, unsupervised, serta semi-supervised learning. Model terbaik ditemukan melalui metode domain adaptive pre-training dengan IndoBERT-large sebagai model acuan dan melibatkan unsupervised learning. Selain itu, model supervised learning juga digunakan dengan menggunakan model konvensional, dan hasilnya digunakan untuk analisis kontribusi dari fitur-fitur yang digunakan dalam klasifikasi. Terakhir, penelitian ini mengevaluasi kembali anotasi yang dilakukan oleh manusia dengan menggunakan kata kunci sebagai pendekatan untuk mengurangi kesalahan dalam dataset. Dengan pendekatan ini, berhasil ditemukan beberapa kesalahan anotasi pada dataset yang dianotasi oleh manusia.

.....

The consultation questions on an online forum need to be answered by the appropriate specialist doctors to provide accurate and beneficial answers to the users asking the questions. In relation to this, this study discusses the development of a model that can automatically direct a health consultation question to a doctor with the corresponding specialization. Furthermore, the constructed model is a multi-label classification model because a question can be associated with more than one specialization. There are several issues addressed in this work. This research begins by evaluating the effectiveness of rule-based mapping methods in predicting data annotated by experts, and the results show a satisfactory level of success. Furthermore, a multi-label classification model is developed to classify the specialist domains of doctors. The model training is performed using various methods, including supervised learning, unsupervised learning, and semi-supervised learning. The best model is found through domain adaptive pre-training using IndoBERT-large as the reference model and involving unsupervised learning. Additionally, the supervised learning model is also used with a conventional model, and the results are used to analyze the contribution of the features used in the classification. Lastly, this research re-evaluates the annotations made by humans using keyword-based approaches to reduce errors in the dataset. With this approach, several annotation errors were successfully identified in the dataset annotated by humans.