

INTISARI

Klasifikasi Pertanyaan dari *Direct Message* dan Komentar Instagram Statistika UII berdasarkan *Latent Dirchlet Allocation* menggunakan *Support Vector Machine*

(Studi Kasus : Informasi tentang Program Studi Statistika Universitas Islam
Indonesia)

Syavira Susriza Fitri

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Saat ini interaksi antara pengguna dan akun media sosial semakin penting dalam lingkungan digital. Di prodi statistika UII akun Instagram menjadi salah satu alat komunikasi dengan menerima pertanyaan-pertanyaan melalui *Direct Message* (DM) dan komentar yang membutuhkan respon cepat dan tepat. Penggunaan Instagram juga sebagai alat promosi dengan memiliki pertanyaan dari DM maupun komentar yang antusias terhadap Prodi Statistika yang penting untuk digali informasinya. Pertanyaan-pertanyaan yang terkumpul kemudian dijadikan sebagai bahan tahapan insisiasi pembuatan chatbot. Oleh karena itu, penelitian ini menggunakan pendekatan LDA untuk mengidentifikasi topik-topik yang umum muncul dalam pertanyaan-pertanyaan yang masuk. Metode ini memungkinkan pengelompokan pertanyaan berdasarkan distribusi kata-kata yang muncul dalam teks. Selanjutnya, SVM digunakan untuk membangun model klasifikasi yang dapat membedakan kategori berdasarkan hasil dari pemodelan topik LDA yaitu PMB, Akademik, Biaya, dan Grup Maba. Dari data pertanyaan yang didapatkan sebanyak 530 pertanyaan. Setelah mendapatkan hasil olah data menggunakan LDA, diperoleh 5 topik salah satunya yaitu Program Magister di Statistika. Kemudian pada klasifikasi, evaluasi didapat dari confusion matrix menunjukkan bahwa model bekerja cukup baik yang didukung dari hasil perhitungan nilai *accuracy* sebesar 87,7%, *precision* sebesar 88%, *recall* sebesar 88%, dan *F1-score* sebesar 87%. Untuk pengujian 20 pertanyaan baru diperoleh hasil menjawab tepat sebanyak 19 pertanyaan sesuai kategori.

Kata Kunci : PMB, Akademik, Biaya, Grup Maba, LDA, SVM.

ABSTRACT

The Classification Questions from Direct Messages and Comments of Statistics UII Instagram's based Latent Dirichlet Allocation using Support Vector Machine

(Case Study : Information about the Statistics Study Program at the Islamic University of Indonesia)

Syavira Susriza Fitri

Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

The interaction between users and social media accounts is increasingly important in a digital environment. In the UII statistics study program, the Instagram account is a means of communication by receiving questions via Direct Message (DM) and comments that require a quick and precise response. The use of Instagram is also a promotional tool by having questions from DMs or enthusiastic comments about statistics study programs which are important for information to be explored. The questions collected are then used as material for the initiation stages of making a chatbot. However, the classification of these questions was processed manually by the UII Statistics team seems to be inefficient with a large number of questions from Instagram account users. Therefore, this study uses the LDA approaches to identify general topics for incoming questions. This method allows grouping questions based on the distribution of words that appear in the text. Furthermore, SVM is used to build a classification model that distinguishes categories based on the results of the LDA topic modeling, namely PMB, Akademik, Biaya, and Group Chat Maba. After getting the results of data processing using LDA, 5 topics were obtained, one of which was the Master's Program in Statistics. Then in classification, the evaluation is obtained from the confusion matrix gives good results supported by the calculation of the accuracy value at 89.6%, precision at 86%, recall at 89%, and f1-score at 87%. Meanwhile, testing 20 new questions and managed to answer about 19 questions according to the category.

Keywords: PMB, Akademik, Biaya, Grup Maba, LDA, SVM.