

INTISARI

KLASIFIKASI ESAI JAWABAN PENDAFTAR IISMA 2022 MENGUNAKAN TEXT MINING

Zul Faza Makarima
19/442493/PA/19242

IISMA atau Indonesian International Student Mobility Awards adalah Salah satu program dari kurikulum Merdeka Belajar - Kampus Merdeka yang mengirimkan mahasiswa dan mahasiswi terbaik bangsa untuk berkuliah di universitas luar negeri selama 1 semester. Dengan meningkatnya pendaftar pada setiap tahunnya, maka beban panitia untuk menyeleksi esai pendaftar pun meningkat. Namun sumber daya manusia dan waktu yang dimiliki panitia sangat terbatas, sehingga pengklasifikasi esai perlu dibuat agar panitia tidak perlu membaca semua esai dan mempercepat proses penyeleksian.

Pada penelitian ini dilakukan klasifikasi esai pendaftar IISMA tahun 2022 menggunakan metode BERT dengan pendekatan 3 label dan 2+2 label dan nilai performa yang dicari adalah akurasi, presisi, *recall*, dan *F1-Score*. Data yang digunakan adalah sejumlah 2304 dengan komposisi yang tidak seimbang. Esai yang digunakan adalah esai dengan topik “*Explain about your health condition (i.e., dental and oral treatment/medical treatment/covid-19 medication/mental health therapy/counseling)!*”.

Hasil dari penelitian ini adalah performa berupa *f1-score* untuk model dengan pendekatan 3 label pada masing-masing kelasnya adalah 80% untuk kelas aman, 70% untuk kelas meragukan, dan 50% untuk kelas berisiko. Serta *f1-score* untuk model dengan pendekatan 2+2 label pada masing-masing kelasnya adalah 88% untuk kelas aman, 70% untuk kelas meragukan, dan 57% untuk kelas berisiko.

Kata Kunci: BERT, Klasifikasi Dokumen Teks

ABSTRACT

CLASSIFICATION OF ANSWERS FOR IISMA 2022 REGISTERS USING TEXT MINING

Zul Faza Makarima
19/442493/PA/19242

IISMA or Indonesian International Student Mobility Awards is one of the programs from the Merdeka Learning curriculum - Merdeka Campus which sends the nation's best students and students to study at foreign universities for 1 semester. With the increase in registrants every year, the committee's burden to select applicants' essays also increases. However, the committee's human resources and time are very limited, so an essay classifier needs to be made so that the committee does not need to read all the essays and speed up the selection process.

In this study, the essay classification of IISMA registrants in 2022 was carried out using the BERT method with a 3 label and 2+2 label approach and the performance values sought were accuracy, precision, recall, and F1-Score. The data used are 2304 with an unbalanced composition. The essay used is an essay with the topic "Explain about your health condition (i.e., dental and oral treatment/medical treatment/covid-19 medication/mental health therapy/counseling)!".

The result of this research is that the performance in the form of f1-score for the model with a 3 label approach in each class is 80% for the safe class, 70% for the doubtful class, and 50% for the risk class. And the f1-score for the model with a 2+2 label approach in each class is 88% for the safe class, 70% for the questionable class, and 57% for the risk class.

Keywords : BERT Text, Document Classification