

Abstrak

Selain Al-Qur'an, Hadis merupakan pedoman hidup yang harus diikuti oleh setiap Muslim di dunia ini. Teknologi untuk mengklasifikasikan teks dan kalimat, termasuk mengkategorikan hadis, berkembang seiring dengan kemajuan zaman. Model yang digunakan untuk melakukan klasifikasi juga telah dikembangkan dan dioptimalkan seperti penggunaan algoritma XGBoost yang lebih optimal dari algoritma tree sebelumnya. Hal ini juga dapat memudahkan kita sebagai umat muslim dalam mempelajari hadis dengan mengkategorikannya berdasarkan anjuran, larangan, dan informasi. Penelitian ini melakukan klasifikasi terjemahan teks hadis berbahasa Indonesia berdasarkan anjuran, larangan, dan informasi dengan menggunakan algoritma XGBoost, TF-IDF untuk ekstraksi fitur, dan Chi-Square untuk seleksi fitur. Pada penelitian ini dilakukan percobaan dengan mengubah urutan proses preprocessing pada bagian stopwords removal dan stemming, melakukan proses klasifikasi dengan dan tanpa menggunakan chi-square sebagai seleksi fitur, serta menambahkan nilai parameter pada saat proses pemodelan dengan XGBoost dan berhasil mendapatkan hasil akhir sebesar 75%, 75%, 74%, dan 74% untuk akurasi, precision, recall, dan F1-Score.

Kata kunci : Teks klasifikasi, Hadist terjemahan bahasa Indonesia, XGBoost, TF-IDF, Chi-Square