

Identifikasi Ayat-Ayat Al-Quran yang Sulit Menggunakan *Random Forest* dan SVM

Satya Rayyis Baruna¹, Kemas Muslim Lhaksana²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹rayyis@students.telkomuniversity.ac.id,

²kemasmuslim@telkomuniversity.ac.id

Abstrak

Al-Qur'an terdiri dari ribuan ayat dengan gaya bahasa Arab yang bervariasi. Beberapa ayat mudah dipahami, namun sebagian lainnya terasa sulit, khususnya bagi pembaca yang tidak memiliki pemahaman bahasa Arab yang kuat. Hal ini menjadi tantangan dalam proses pembelajaran dan pemahaman Al-Qur'an, khususnya bagi pemula. Sayangnya, belum banyak tersedia sistem yang dapat secara otomatis mengidentifikasi ayat-ayat yang sulit berdasarkan karakteristik bahasanya. Penelitian ini merancang sistem klasifikasi otomatis untuk mengenali tingkat kesulitan ayat, dengan menggunakan 625 ayat yang telah dilabeli "mudah" dan "sulit" oleh responden. Algoritma *Random Forest* digunakan sebagai metode utama, dan SVM sebagai pembanding untuk melihat perbedaan kinerja antara dua pendekatan yang berbeda. Teks ayat diproses melalui tahap *preprocessing*, transformasi fitur menggunakan TF-IDF, serta penyeimbangan data dengan RUS dan SMOTE. Hasil pengujian menunjukkan bahwa setelah dilakukan penyeimbangan data, model *Random Forest* mampu meningkatkan *recall* untuk ayat sulit dari 14,89% menjadi 63,83% dan *F1-score* dari 22,22% menjadi 51,28%, dengan akurasi keseluruhan sebesar 63,69%. Meskipun terjadi sedikit penurunan akurasi secara keseluruhan, model justru menjadi lebih seimbang dalam mengenali kedua jenis ayat. Penelitian ini berkontribusi dalam pengembangan sistem klasifikasi ayat sulit, yang dapat dimanfaatkan untuk mendukung pembelajaran Al-Qur'an secara lebih adaptif dan terarah.

Kata kunci: al-qur'an, pembelajaran mesin, klasifikasi teks, *random forest*, TF-IDF, ayat sulit
