

1. Pendahuluan

Latar Belakang

Al-Qur'an adalah kitab suci umat Islam yang diturunkan kepada Nabi Muhammad SAW sebagai pedoman serta sumber pengajaran dan hukum bagi manusia. Al-Qur'an terdiri dari 6236 ayat yang dibagi menjadi 114 surah ke dalam 30 juz. Sebagai *Kalamullah* (firman Allah), Al-Qur'an memiliki struktur kalimat yang unik yang tidak dimiliki oleh kitab lainnya. Al-Qur'an juga memiliki banyak karakteristik yang telah diteliti oleh para ulama Islam di dunia. Penelitian tersebut difokuskan pada ciri khusus Al-Qur'an dalam beberapa aspek. Salah satunya adalah aspek kandungan dari setiap ayat Al-Qur'an yang menjadi sumber pengajaran utama dari agama Islam.

Dalam hal memahami Al-Qur'an, terdapat beberapa hal yang perlu dilakukan. Pertama, menerjemahkan Al-Qur'an ke dalam bahasa yang dimengerti. Salah satu bahasa terjemahan Al-Qur'an yang paling populer adalah terjemahan bahasa Inggris, oleh karena bahasa Inggris disepakati sebagai bahasa Internasional. Selain itu, untuk memahami Al-Qur'an, perlu mengetahui topik yang dibahas dalam setiap ayat. Di era teknologi ini, ayat Al-Qur'an dapat dikategorikan ke dalam satu atau lebih ayat secara otomatis dengan bantuan algoritma *Machine Learning*. Dalam *machine learning*, sistem ini disebut dengan *Multilabel Text Classification* dimana tujuan utamanya adalah untuk mengklasifikasikan suatu teks ke satu atau lebih topik [11].

Klasifikasi teks dilakukan dengan mengubah kata-kata menjadi angka yang direpresentasikan dalam bentuk *Vector Space Model* yang digunakan pada tahap ekstraksi fitur agar dapat diproses oleh algoritma *machine learning* [10]. Metode pembobotan kata yang paling sering digunakan adalah TF-IDF [15]. Namun, metode *word centrality* untuk pembobotan kata menjadi salah satu metode yang memberikan hasil yang memuaskan di beberapa penelitian [9][15].

Topik dan Batasannya

Untuk mengukur *word centrality* dan membuat model klasifikasi, dataset yang digunakan adalah data ayat Al-Qur'an yang telah diberi label dalam 10 topik. Penelitian ini dibangun dengan beberapa tahap preprocessing yakni *text cleaning*, *tokenization*, dan *stopword removal*. Tahap ekstraksi fitur dibangun dengan dua metode yakni TF-IDF dan TW-IDF (dengan nilai *word centrality*) untuk dilakukan perbandingan. Tahap pembuatan model klasifikasi dibangun dengan hanya dibatasi pada empat algoritma *machine learning* yakni SVM, *Naïve Bayes*, KNN, dan *Decision Tree*.

Tujuan

Tujuan dari penelitian ini adalah membangun sistem klasifikasi Topik pada Ayat Al-Qur'an dalam terjemahan Bahasa Inggris dengan menambahkan hasil pengukuran *Word Centrality* sebagai salah satu metode *Feature Extraction* menggunakan perbandingan empat algoritma klasifikasi yakni *Support Vector Machine (SVM)*, *Naïve Bayes*, *K-Nearest Neighbor (K-NN)* dan *Decision Tree*. Serta untuk membandingkan hasil pengujian sistem terhadap 2 (dua) skenario pengujian yakni menggunakan *Stopword Removal* dan tanpa menggunakan *Stopword Removal*.

Organisasi Tulisan

Penulisan artikel ini tersusun sebagai berikut. Bagian 1 menjelaskan pendahuluan, Bagian 2 menjelaskan perihal studi terkait, Bagian 3 menjelaskan rancangan sistem dan proses penelitian, Bagian 4 menjelaskan analisis hasil penelitian, dan kesimpulan penelitian dijelaskan pada Bagian 5.