

Abstrak

Informasi yang tersebar melalui internet banyak dimanfaatkan oleh orang-orang untuk mencari hal apa saja. Salah satunya yang banyak dicari di internet adalah informasi terkait pengetahuan agama Islam. Namun, banyaknya informasi yang tersedia dari berbagai macam sumber sehingga sulit bagi orang-orang untuk menemukan informasi yang tepat. Penelitian terkait topik ini sudah pernah dilakukan oleh peneliti sebelumnya, namun dataset yang digunakan hanya berasal dari satu sumber saja. Sehingga, pada penelitian ini dibangun sistem klasifikasi topik tanya jawab Islam menggunakan metode *Naïve Bayes* dan TF-IDF. Pada penelitian ini menggunakan 1000 data artikel tanya jawab yang diambil dari situs *web* konsultasi Islam yaitu *rumahfiqih.com* dan *islamqa.info*. Klasifikasi *multi-class* menggunakan lima kategori yang dilabelkan secara manual menggunakan kelas kategori pada website tersebut. Dengan beberapa skenario pengujian, pada penelitian ini metode klasifikasi *Naïve Bayes* dengan menggunakan TF-IDF (level n-gram) dengan *max features* sebanyak 1000 pada rasio data splitting 70:30 menghasilkan akurasi tertinggi sebesar 81%. Nilai akurasi sebesar 81% juga dihasilkan oleh metode klasifikasi SVM, namun perbedaannya pada SVM nilai akurasi tertinggi menggunakan TF-IDF (level kata). Diharapkan pada penelitian selanjutnya akan digunakan lebih banyak sumber situs *web* dan penggunaan metode klasifikasi dan *feature extraction* lain dengan nilai yang lebih optimal dari penelitian sebelumnya.

Kata Kunci: tanya jawab, klasifikasi topik, multi-class, naïve bayes, tf-idf