

## 1. Pendahuluan

Hadits adalah segala perkataan (sabda), perbuatan dan ketetapan atau persetujuan dari Nabi Muhammad SAW yang dijadikan ketetapan ataupun hukum dalam agama Islam. Hadits dijadikan sumber hukum dalam agama Islam selain Al-Qur'an, Ijma dan Qiyas. Kedudukan hadits merupakan sumber hukum kedua setelah Al-Qur'an[1]. Pada dewasa ini teknologi informasi yang terus berkembang semakin mempermudah masyarakat dalam mendapatkan informasi, terutama dengan menggunakan internet. Informasi yang didapatkan di internet lebih banyak informasi berupa dokumen teks. Hal ini berpengaruh terhadap bertambahnya jumlah hadits shahih Bukhari terjemahan Bahasa Indonesia berupa dokumen teks. Bagi masyarakat yang tidak memiliki cukup waktu, mengakibatkan kesulitan untuk memahami setiap makna dari setiap hadits yang disebabkan jumlah hadits yang sangat banyak. Jumlah dokumen yang sangat banyak mendorong peneliti untuk melakukan penelitian dalam membuat sistem pengolahan terhadap kumpulan dokumen hadits tersebut kedalam kelas topik larangan, anjuran, atau informasi. Mengelola dokumen teks dengan jumlah yang banyak merupakan hal yang sulit untuk dilakukan secara manual. Oleh karena itu, penelitian ini menggunakan teknik yang dapat mengkategorikan kelas dari dokumen secara otomatis. Klasifikasi merupakan salah satu teknik dalam *text mining* dan *data mining* yang digunakan dalam proses mencari model atau fungsi yang menjelaskan atau membedakan kelas-kelas pada data dan konsep yang bertujuan untuk menggunakan model tersebut dalam melakukan prediksi terhadap *data testing*. Salah satu metode klasifikasi adalah *Support Vector Machine* (SVM) yang memiliki kelebihan yaitu mampu memberikan performansi yang baik dalam dataset yang memiliki fitur yang sangat banyak. *Latent Semantic Indexing* (LSI) sebagai metode pemilihan fitur digunakan untuk mereduksi fitur dengan menghilangkan fitur yang tidak penting dari data (*noise term*). Terdapat dua metode untuk meningkatkan tingkat akurasi prediksi dan model dalam klasifikasi yaitu *Boosting* dan *Bootstrap Agregating* (*bagging*). Sistem klasifikasi pada penelitian ini akan memanfaatkan metode *bagging* dengan tujuan untuk meningkatkan nilai akurasi.

Terdapat beberapa batasan masalah dalam penelitian ini adalah *dataset* yang digunakan adalah data hadits shahih Bukhari terjemahan Bahasa Indonesia dengan jumlah 1650 hadits dengan komposisi sebanyak 500 untuk setiap label informasi, larangan dan anjuran sebagai *data training*, dan sebanyak 50 untuk setiap label informasi, larangan dan anjuran sebagai *data testing*. Dilakukan penghilangan *stopwords* pada proses *preprocessing* dataset.

Pada jurnal ini, disusun sebagai berikut. Pada bagian kedua, terdapat studi terkait yang menjelaskan mengenai metode penelitian sebelumnya yang berkaitan dengan metode penelitian yang diterapkan pada penelitian yang sedang dikerjakan ini. Pada bagian ketiga, menjelaskan tentang sistem klasifikasi yang dibangun dan diimplementasikan pada penelitian ini dengan judul pengklasifikasian topik hadits shahih Bukhari terjemahan Bahasa Indonesia menggunakan *Latent Semantic Indexing* dan *Support Vector Machine*. Pada bagian keempat, terdapat evaluasi dan analisis hasil pengujian yang telah didapatkan. Kemudian pada bagian kelima, menjelaskan kesimpulan yang didapatkan pada penelitian ini.