

BAB I PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi yang terus berkembang semakin mempermudah masyarakat dalam mendapatkan berbagai informasi khususnya di media *online*. Sebagian besar dari informasi yang didapatkan di internet adalah berupa data teks. Pertumbuhan data teks yang semakin pesat mendorong para peneliti membangun sebuah sistem untuk penanganan dan pengorganisasiannya untuk mendapatkan informasi baru yang dapat dimanfaatkan dari kumpulan data teks tersebut. Salah satu teknik yang sering digunakan adalah klasifikasi teks. Tujuan dari teknik ini adalah mengkategorikan dokumen atau teks ke dalam beberapa kelas. Teknik ini sudah banyak digunakan pada banyak penelitian dengan berbagai domain, namun masih sedikit yang melakukan penelitian dengan mengangkat domain teks kitab agama, salah satunya adalah Hadits. Hadits merupakan segala sabda, perbuatan, *taqirir*, dan hal-ihwal yang berasal dari Nabi Muhammad Shallallahu ‘Alaihi Wasallam, sehingga hadits juga merupakan sumber hukum kedua setelah Al-Qur’an yang dijadikan sebagai pedoman oleh umat Muslim [1] [2].

Bagi umat muslim, mengikuti dan mempraktekkan segala perbuatan Nabi Muhammad Shallallahu ‘Alaihi Wasallam yang tertuang pada hadits dalam kehidupan sehari-hari merupakan salah satu *sunnah*. Namun, bagi orang awam ataupun orang muslim yang ingin belajar tentang hadits dan tidak memiliki waktu yang cukup, mungkin masih sedikit kesulitan untuk memahami satu persatu makna dari setiap hadits, mengingat jumlah hadits yang cukup banyak. Menurut aplikasi Lidwa jumlah hadits Shahih Bukhari berjumlah kurang lebih 7.008 hadits termasuk hadits yang berulang [3].

Pada dasarnya, hadits memiliki klasifikasi yang didasarkan pada kredibilitas sebuah hadits dapat dijadikan sebagai dasar hukum atau tidak, salah satunya adalah hadits shahih. Hadits shahih merupakan hadits yang paling kuat jalan periwayatannya atau *sanadnya* dari Nabi Muhammad Shallallahu ‘Alaihi Wasallam dan dianggap sebagai hadits yang boleh dijadikan sebagai dasar hukum. Salah satu

hadits populer yang kumpulan haditsnya dijadikan sebagai acuan periwayat hadits lainnya adalah hadits shahih yang diriwayatkan oleh Imam Bukhari [4]. Disamping hadits yang dapat diklasifikasikan berdasarkan kredibilitasnya, apabila dilihat dari isi hadits ternyata terdapat makna atau informasi yang dapat diambil. Informasi tersebut dapat berupa sebuah perintah berupa anjuran, larangan, atau informasi berupa cerita dari Rasulullah shallallahu ‘alaihi wasallam terkait segala perilaku dan perkataan beliau dalam menyelesaikan berbagai masalah.

Berdasarkan permasalahan diatas, maka salah satu solusi yang dapat dilakukan adalah dengan membangun sebuah sistem untuk mengkategorikan data hadits kedalam kelas anjuran, larangan dan informasi secara otomatis.

Dalam proses pengklasifikasian hadits ini, penulis menemukan beberapa tantangan yaitu yang pertama adalah data hadits yang digunakan ini diambil dari aplikasi kumpulan Hadits Lidwa yang belum memiliki label dan masih mengandung *sanad*, sehingga untuk proses pelabelan penulis melakukan pemisahan *sanad* secara manual dan melakukan proses *hand labeling* untuk menentukan kelas anjuran, larangan dan informasi berdasarkan fitur-fitur unik yang ada pada setiap kelas. Kemudian, tantangan lain yang dihadapi adalah adanya kalimat ambigu yang ada pada hadits, contohnya terdapat hadits yang merupakan kelas anjuran, namun juga memiliki fitur yang merepresentasikan kelas larangan. Selain itu, terdapat beberapa hadits yang ada di dalam kelas informasi mengandung fitur-fitur yang seharusnya ada pada kelas anjuran dan larangan. Hal itu mengakibatkan kesalahan pelabelan yang dilakukan oleh *classifier*. Oleh karena itu, *simple rule-based system* berdasarkan kemunculan fitur “-lah” dan beberapa kata penting lainnya tidak akan bekerja dengan maksimal. Untuk itu, penelitian ini memilih menggunakan teknik *machine learning* yang salah satunya adalah metode *Support Vector Machine*. Pemilihan metode *Support Vector Machine* sebagai *classifier* pada penelitian klasifikasi hadits ini adalah karena metode ini memiliki strategi untuk meminimalkan rata-rata *error* pada data *training* dan strategi SRM (*Structural Risk Minimization*) untuk menangani klasifikasi data yang tidak terdapat dalam data *training* [5].

Terdapat beberapa penelitian klasifikasi teks Bahasa Indonesia yang menggunakan metode *Support Vector Machine*, salah satunya adalah penelitian

yang dilakukan oleh Aqsath Rasyid Naradhipa dan Ayu Purwarianti. Klasifikasi teks yang dilakukan pada penelitian tersebut adalah klasifikasi sentimen pada pesan yang ada di media sosial. Peneliti membandingkan antara *classifier* yang dibangun dengan menggunakan *Maxmum Entropy* dan *Support Vector Machine*. Akurasi optimal dari penelitian tersebut didapat dari *classifier* yang dibangun dengan *Support Vector Machine* yaitu sebesar 86,66%. [6]

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah di uraikan diatas, terdapat beberapa pertanyaan dalam penelitian ini, antara lain :

1. Bagaimana cara membangun *text classification* pada data Hadits Shahih Bukhari ?
2. Bagaimana menganalisis model dan performansi dari sistem *text classification* pada data hadits Shahih Bukhari ?

1.3 Tujuan

Berdasarkan latar belakang yang telah diuraikan diatas, terdapat beberapa tujuan yang hendak di capai pada penelitian ini, antara lain :

1. Membangun sistem *text classification* yang dapat mengklasifikasikan data Hadits Shahih Bukhari menggunakan metode *Support Vector Machine*.
2. Menganalisis model dan performansi dari sistem *text classification* pada data Hadits Shahih Bukhari menggunakan *classifier Support Vector Machine*.

1.4 Batasan Masalah

Batasan masalah dari tugas akhir ini adalah sebagai berikut :

1. Data hadits yang digunakan berasal dari aplikasi Lidwa dan merupakan hadits terjemahan Bahasa Indonesia.
2. Kelas hadits dibagi menjadi tiga yaitu anjuran, larangan, dan informasi.
3. Untuk data hadits yang mengandung fitur kelas anjuran dan larangan maka akan di masukkan ke dalam kelas larangan.
4. Data hadits yang digunakan adalah 1651 data.

1.5 Metodologi Penyelesaian Masalah

Rencana kegiatan pada penelitian ini terdiri dari beberapa tahapan sebagai berikut:

1. Identifikasi Masalah

Tahap ini merupakan tahap awal dari penelitian *text classification* pada Hadits Shahih Bukhari yaitu dimulai dengan mencari data hadits yang valid untuk digunakan pada penelitian tugas akhir.

2. Studi literatur dan pengumpulan data

Pada tahap ini, peneliti melakukan kajian pada paper dan jurnal yang berkaitan dengan tugas akhir ini serta melakukan pengumpulan dataset dengan cara melakukan proses *crawling* dari app.lidwa.com dan melakukan pelabelan secara manual.

3. Perancangan Sistem

Pada tahap ini, peneliti membuat rancangan sistem berupa *flowchart*, merancang skenario pengujian serta membuat desain sistem.

4. Penulisan Proposal Tugas Akhir

Pada tahap ini, peneliti menulis rencana dan skenario yang akan diimplementasikan pada sistem kedalam sebuah laporan proposal untuk memudahkan peneliti dalam proses penelitian..

5. Implementasi

Pada tahap ini peneliti mulai melakukan pembangunan sistem berdasarkan rancangan pada tahap sebelumnya.

6. Pengujian

Pada tahap ini peneliti melakukan pengujian sistem yang sudah dibuat dengan data yang dimiliki. Peneliti melakukan pengujian dengan mengatur parameter SVM yaitu *degree*, *gamma*, dan *C* pada kernel "*polynomial*" dan "*RBF*".

7. Analisis

Pada tahap ini peneliti menganalisis semua hasil dari tiap proses yang ada pada sistem serta hasil akhir dari pengujian sistem.

8. Penulisan Laporan Akhir

Pada tahap ini peneliti mendokumentasikan hasil yang telah dilakukan ke dalam bentuk laporan akhir sebagai persyaratan sidang dan digunakan untuk penelitian lebih lanjut oleh pihak lain.

1.6 Sistematika Penulisan

Untuk membantu pembaca dalam memahami laporan tugas akhir ini, maka penulis mengelompokkan laporan menjadi beberapa subbab dengan sistematika penulisan sebagai berikut.

1. **Pendahuluan**

Bab ini berisi tentang informasi umum yaitu latar belakang, perumusan masalah, batasan masalah, tujuan, metodologi penyelesaian masalah, dan sistematika penulisan.

2. **Landasan Teori**

Bab ini berisi tentang teori yang digunakan dalam penelitian yang diambil dari literatur terkait.

3. **Perancangan Sistem**

Bab ini berisi tentang gambaran umum sistem dan tahapan tiap prosesnya dan menjelaskan spesifikasi lingkup operasional yang digunakan pada penelitian.

4. **Pengujian dan Analisis**

Bab ini berisi tentang analisis dari hasil penelitian yang telah dilakukan.

5. **Kesimpulan dan Saran**

Bab ini berisi kesimpulan dan saran yang berkaitan dengan hasil analisis sistem.