

# 1. Pendahuluan

## 1.1. Latar Belakang

Hadits adalah segala sesuatu yang dinisbatkan kepada Nabi Muhammad SAW baik berupa perkataan, perbuatan, taqir (sikap diam setuju) dsb. Hadits merupakan sumber hukum tersendiri bagi umat muslim yang tidak dijelaskan dalam Al Qur'an[8] Setiap Hadist terdiri dari 2 bagian yaitu Sanad dan Matan. Sanad adalah untaian nama para penyampai Hadist yang menjamin keaslian dari Hadist itu. Matan sendiri adalah konten dari hadist itu. Setiap hadist diawali dengan Sanad [16].

Umumnya hadits di koleksi oleh beberapa imam besar, salah satunya koleksi hadits yang disusun oleh Imam Bukhari (nama lengkap: Abu Abdullah Muhammad bin Ismail bin Ibrahim bin al-Mughirah al-Ju'fi) yang hidup antara 194 hingga 256 hijriah.

Sebagai sumber hukum tersendiri, maka umat islam dianjurkan untuk mempelajari dan mengamalkannya. Hal tersebut dimudahkan dengan tersedianya cetakan buku kumpulan hadist hingga versi digital. Sejauh ini sumber – sumber tersebut hanya menyediakan hadist berdasarkan kitab pembahasannya namun hingga saat ini belum ada penelitian maupun inovasi pengklasifikasian hadits berdasarkan anjuran, larangan dan yang hanya sekedar informasi. Dengan adanya klasifikasi hadits akan memudahkan masyarakat yang akan mempelajari hadits berdasarkan ketegorinya.

Penelitian mengenai analisis klasifikasi pada kenyatannya telah banyak dilakukan. *Naïve Bayes Classifier* (NBC) ialah salah satu teknik pembelajaran mesin yang cukup sering digunakan untuk menangani hal tersebut. Seperti pada penelitian *Twitter Sentiment Classification* oleh Alec Go, dkk (2009) yang menggunakan *Naïve Bayes* dan dua metode lainnya yaitu SVM dan MaxEnt menghasilkan nilai rata - rata akurasi terbaik dari NBC yaitu 81.375%, selanjutnya penelitian oleh Joko Samodra, dkk (2009) pada Klasifikasi Dokumen Teks Berbahasa Indonesia Menggunakan *Naïve Bayes* juga menunjukkan metode NBC terbukti dapat

memberikan hasil yang baik dalam melakukan klasifikasi dokumen teks berbahasa Indonesia, yang terlihat dari akurasi yang terus meningkat hingga 87,63%.

Namun pada penerapannya Naïve Bayes Classifier merupakan klasifikasi yang mengasumsikan keberadaan atribut suatu kelas tidak terkait atau tidak akan mempengaruhi atribut di kelas lain, atau dikenal dengan *class conditional independence*. Sehingga akan menjadi celah untuk mengurangi keefektifan metode ini dan akibatnya meloloskan dokumen ke dalam kelas tertentu yang bukan kelas seharusnya.

Permasalahan yang sering muncul pada penelitian terkait adalah banyaknya fitur / kata dalam satu dokumen. Hal ini tentu mengganggu akan sangat mempengaruhi proses dari klasifikasi itu sendiri. Oleh karena itu, untuk mengurangi hal tersebut data harus dilakukan beberapa tahapan. Dalam hal tersebut biasanya metode yang dapat digunakan adalah *feature selection*. *Feature selection* adalah suatu proses untuk menyeleksi subset fitur dari sekumpulan fitur asli yang dapat mengurangi jumlah fitur, mempercepat proses algoritma data mining serta memperbaiki performansi data mining dengan cara menghilangkan fitur yang tidak relevan, redundan, noise. [12]. Selain itu dalam proses data mining, tahap *preprocessing* mengambil peranan penting terhadap kualitas data. Seperti *stemming* yang merupakan salah satu tahap dalam *preprocessing* yang bertujuan untuk mengubah suatu kata berimbuhan menjadi kata dasar. Pemotongan imbuhan tersebut akan berdampak pada perhitungan kemunculan kata yang selanjutnya dapat mempengaruhi perhitungan probabilitas suatu kalimat.

Oleh sebab itu pada penelitian ini akan dibahas mengenai hasil akurasi klasifikasi hadits Shahih Bukhari menggunakan metode klasifikasi Naïve Bayes dengan menggunakan *feature selection Chi-Square* serta bagaimana pengaruh tahap *preprocessing stemming* dalam proses klasifikasi data.

## **1.2. Perumusan Masalah**

Berdasarkan latar belakang yang ada, didapatkan masalah yang akan dibahas pada penelitian Tugas Akhir ini adalah :

1. Bagaimana analisis hasil akurasi metode *Naïve Bayes* dalam mengklasifikasi data Hadits?
2. Bagaimana pengaruh *feature selection* menggunakan *Chi-square* dalam mengklasifikasi data hadits?
3. Bagaimana pengaruh penggunaan *stemming* dalam mengklasifikasi data hadits?
4. Bagaimana analisis performansi metode *Naïve Bayes* dalam mengklasifikasi data Hadits?

Adapun batasan masalah yang dilakukan agar pengerjaan Tugas Akhir ini tidak terlalu luas adalah:

1. Data yang digunakan merupakan data Hadits Shahih Bukhari sebanyak 2556 data.
2. Penentuan kelas berupa anjuran, larangan, dan informasi dilakukan atas saran narasumber yang ahli pada bidangnya.
3. Hanya menangani klasifikasi hadits yang masuk ke salah satu kategori saja (anjuran , larangan, dan informasi).
4. Tidak menggunakan Teknik untuk menangani ketidakseimbangan data pada suatu kelas.

### **1.3. Tujuan**

Berdasarkan rumusan masalah yang ada, maka tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui hasil akurasi metode *Naïve Bayes Classifier* dalam mengklasifikasi data Hadits.
2. Mengetahui pengaruh *feature selection* menggunakan *Chi-square* dalam klasifikasi data Hadits.
3. Mengetahui pengaruh *stemming* dalam klasifikasi data Hadits.
4. Mengetahui hasil performansi metode *Naïve Bayes* dalam mengklasifikasi data Hadits.

## **1.4. Metode Penyelesaian**

Metodologi yang digunakan dalam penelitian Tugas Akhir ini meliputi studi literatur, perancangan dan implementasi, pencarian dataset dan dokumentasi.

Penjelasan dari metodologi yang dipakai adalah sebagai berikut:

1. Studi Literatur  
Mempelajari metode dan masalah dari referensi, meliputi metode *Naïve Bayes Classifier*, *Chi-square*, dan *Preprocessing*.
2. Pengolahan Dataset  
Mengumpulkan dan memberi label pada dataset yang akan digunakan dalam penelitian tersebut melalui *software* hadits Sembilan imam.
3. Perancangan dan implementasi  
Membangun sistem yang diimplementasikan dengan bahasa pemrograman.
4. Pengujian dan analisis  
Melakukan pengujian terhadap sistem yang telah dirancang untuk melihat nilai akurasi, performansi dan analisis pengaruh dari metode yang digunakan.
5. Penarikan kesimpulan dan dokumentasi  
Menarik kesimpulan dari hasil penelitian yang telah dilakukan dan menulis laporan dari hasil penelitian dalam bentuk buku tugas akhir.

## **1.5. Sistematika Penulisan**

Penulisan tugas akhir ini dibagi menjadi lima bab yaitu:

1. Bab 1 : Pendahuluan  
Pada Bab ini membahas latar belakang, tujuan, batasan masalah, metodologi penyelesaian masalah, serta sistematika penulisan dari tugas akhir yang dibangun.
2. Bab 2 : Studi Literatur  
Pada Bab ini membahas teori-teori yang terkait dengan pengerjaan tugas akhir. Teori-teori yang dibahas yaitu Klasifikasi, Hadits, *Text Mining*, *Preprocessing*, *Feature Selection*, dan *Naïve Bayes*.
3. Bab 3 : Perancangan Sistem

Pada Bab ini membahas Perancangan Sistem yang digunakan untuk membangun sistem yang dapat mengklasifikasikan hadits berupa anjuran, larangan, dan informasi dan menggunakan metode *Naïve bayes*. Selain itu pada Bab ini juga membahas alur dari setiap Proses yang dilalui beserta contoh dari proses yang dilalui dan Skenario Pengujian yang digunakan pada Bab 4 yaitu Bab Pengujian dan Analisis.

4. Bab 4 : Pengujian dan Analisis

Pada Bab ini membahas hasil pengujian berdasarkan skenario pengujian yang dituliskan pada bab Perancangan Sistem. Selain itu, pada bab ini juga dijelaskan analisis terhadap hasil pengujian tersebut. Hasil dari kegiatan analisis ini menjadi dasar pengambilan kesimpulan.

5. Bab 5 : Kesimpulan dan Saran

Pada Bab ini berisi tentang kesimpulan beserta saran dari keseluruhan hasil pengerjaan Tugas Akhir .