

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi pada zaman modern ini telah sampai kepada era elektronik, ditandai dengan semakin digunakannya teknologi berupa komputer dan jaringan internet sebagai sarana utama penyampaian informasi. Dokumen yang beredar di dunia maya terus tumbuh dan mungkin menjadi kurang efektif dalam pencarian dan penyajian informasi. Seiring dengan kemajuan ilmu pengetahuan dan teknologi mayoritas orang lebih memanfaatkan teknologi sebagai media untuk belajar dan mencari informasi. Seperti halnya dalam mencari makna atau topik dari terjemahan Al-Qur'an.

Al-Quran adalah kitab suci yang merupakan pedoman hidup dan dasar setiap langkah hidup umat manusia. Al-Quran bukan sekedar mengatur hubungan manusia dengan Allah SWT, tetapi juga mengatur hubungan manusia dengan manusia lainnya bahkan dengan alam sekitarnya. Klasifikasi ayat Al-Quran perlu diketahui agar dapat memudahkan dalam memahami makna dari ayat-ayat Al-Quran. Al-Quran dapat diklasifikasikan ke dalam berbagai klasifikasi seperti klasifikasi berdasarkan susunan surat, berdasarkan tempat diturunkannya ayat Al-Quran atau berdasarkan pembahasannya dan lain-lain [1].

Banyaknya data teks dari terjemahan ayat Al-Quran menjadi masalah ketika data tersebut masih tidak terstruktur. Untuk itu dibutuhkan metode *text mining* untuk mengubah data data yang tidak terstruktur menjadi lebih terstruktur [2]. *Text mining* merupakan teknik penggalian data berupa teks, data yang digali biasanya berasal dari

suatu dokumen. Tujuan dari dilakukannya text mining ini adalah untuk mencari kata-kata yang dapat mewakili isi dokumen, sehingga dengan dilakukannya teknik ini dapat mengetahui pembahasan inti dari suatu dokumen [1]. Selain itu *text mining* juga digunakan untuk mengelompokan suatu dokumen ke dalam kategori-kategori tertentu berdasarkan konten yang dimuatnya.

Text classification merupakan sub bidang dari *text mining* yang menentukan kategori dari sebuah dokumen berdasarkan kontennya [2]. Metode yang sering digunakan untuk melakukan *text classification* antara lain adalah metode *Naïve Bayes*, *Support Vector Machine (SVM)*, *Random Forest*, *Decision Tree*, dan yang lainnya. *Naïve Bayes* sangat luas dipakai dalam berbagai bidang khususnya untuk *text classification*. *Naïve Bayes* merupakan metode yang termasuk ke dalam *Supervised Learning* yang membutuhkan data pelatihan yang telah diberikan label terlebih dahulu [3].

Dalam Al-Qur'an terjemahan versi Al-Hadi melampirkan bahwa ayat-ayat Al-Qur'an terklasifikasi menjadi delapan 8 kategori yaitu: Aqidah, Syariah, Akhlak, Ilmu-Ilmu, Kisah, Alam Dunia, Alam Ghaib, dan Alam Akhirat. Namun belum semua ayat memiliki kategori. Dalam penelitian ini ayat-ayat tersebut akan dikategorikan dengan menggali informasi dari ayat-ayat yang sudah memiliki kategori atau label. Selain itu terdapat beberapa ayat yang memiliki lebih dari 1 kategori seperti pada Q.S. Al-Baqarah ayat 2 yang memiliki 4 kategori yaitu Aqidah, Syariah, Kisah dan Alam Dunia. Fakta ini menunjukkan bahwa klasifikasi ayat Al-Qur'an berbeda dengan klasifikasi pada umumnya, dimana

setiap data/dokumen hanya tergolong ke dalam sebuah kelas. Kasus klasifikasi seperti ini dapat disebut sebagai klasifikasi *multi-label* [4].

Dalam kasus klasifikasi multi-label maka perlu dilakukan perubahan dari bentuk dasar formula *Naïve Bayes* agar bisa mengklasifikasikan suatu dokumen ke banyak kelas. *Multinomial Naïve Bayes* merupakan turunan dari *Naïve Bayes* yang denominatornya dianggap konstan [5]. Berdasarkan laporan dari hasil penelitian sebelumnya metode *Naïve Bayes* memberikan hasil yang bagus ketika diimplementasikan baik ke dalam dokumen berbahasa Inggris maupun dokumen berbahasa Indonesia [5].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka permasalahan yang dapat dirumuskan adalah "Bagaimana melakukan klasifikasi multi-label terhadap terjemahan ayat Al-Qur'an berbahasa Indonesia dengan metode *Multinomial Naïve Bayes*?"

1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Data terjemahan bahasa Indonesia Al-Quran yang digunakan dalam penelitian ini diambil dari *qurandatabase.org*.

2. Terdapat 8 topik klasifikasi berdasarkan Al-Quran terjemahan versi Al-Hadi yaitu Aqidah, Akhlak, Syariah, Ilmu, Kisah, Alam Dunia, Alam Ghaib, dan Alam Akhirat.
3. Algoritma yang digunakan dalam penelitian ini adalah *Multinomial Naïve Bayes*.

1.4 Maksud Dan Tujuan Penelitian

Berdasarkan permasalahan di atas, tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Untuk mengklasifikasi terjemahan ayat Al-Quran dengan menggunakan algoritma *Multinomial Naïve Bayes*.
2. Untuk mengetahui performa algoritma *Multinomial Naïve Bayes* dalam mengklasifikasikan terjemahan ayat Al-Quran ke dalam 8 topik yang telah ditentukan.
3. Untuk memberikan referensi bagi peneliti lain tentang algoritma *Multinomial Naïve Bayes* dalam proses klasifikasi ayat Al-Quran terjemahan bahasa Indonesia.

1.5 Metodologi Penelitian

Dalam melakukan penelitian ini diperlukan beberapa metode agar mencapai hasil yang diharapkan yaitu sebagai berikut:

1.5.1 Metode Pengumpulan Data

Dalam memperoleh data dan informasi sebagai penunjang penelitian ini dilakukan dengan beberapa metode:

1.5.1.1 Metode Studi Pustaka

Metode pencarian data dari buku, jurnal, artikel, dan internet atau referensi lainnya yang merupakan penunjang dalam memperoleh data dan informasi untuk melengkapi penyusunan laporan yang berhubungan dengan masalah yang dibahas.

1.5.1.2 Metode Scraping

Metode yang digunakan untuk memperoleh *dataset* yang digunakan dalam penelitian ini adalah dengan melakukan *web-scraping* dari website *alquranahadi.com*.

1.5.2 Metode Eksperimental

Penelitian ini merupakan peneliti eksperimental yaitu melakukan uji coba terhadap algoritma *Multinomial Naïve Bayes* untuk melakukan klasifikasi terhadap terjemahan ayat Al-Qur'an Bahasa Indonesia.

1.5.3 Metode Evaluasi

Dari hasil pengujian akan dilakukan evaluasi dengan menghitung nilai *Hamming Loss* dan digambarkan ke dalam bentuk tabel dan grafik. Semakin kecil nilai *Hamming Loss* maka semakin baik performa dari model klasifikasi yang dibuat [6].

1.6 Sistematika Penulisan

Guna memudahkan pembaca dalam mengikuti apa yang dipaparkan dalam laporan skripsi ini, sistematika penulisan tersusun dari 5 (lima) bab yaitu:

BAB I PENDAHULUAN

Bab ini berisikan gambaran umum penelitian yang dilakukan, meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian metode penelitian, dan sistematika penelitian.

BAB II LANDASAN TEORI

Bab ini membahas tentang dasar-dasar teori yang mendukung dalam melakukan proses pengklasifikasian ayat Al-Quran menggunakan algoritma *Multinomial Naïve Bayes* mulai dari tahap *data processing*, *stopword removal*, *feature extraction*, *classification* dan evaluasi.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tentang metodologi yang digunakan dalam penelitian yang mencakup metode klasifikasi teks menggunakan algoritma *Multinomial Naïve Bayes*, serta metode-metode penunjang lainnya seperti pengumpulan dataset, *pre-processing*, *Feature Selection*, *Feature extraction*, *Classification Stage*, dan Evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi penjelasan dari hasil dan pembahasan mengenai klasifikasi pada terjemahan Al-Qur'an. Hasil klasifikasi yang didapatkan menggunakan algoritma *Multinomial Naïve Bayes* berdasarkan 4 tahap proses klasifikasi *pre-processing*, *Feature Selection*, *Feature extraction*, *Clasification Stage*. Selanjutnya pembahasan dilakukan dengan melihat pada teori sebelumnya dengan memperhatikan dan menimbang berdasarkan eksperimen yang dilakukan.

BAB V PENUTUP

Bab ini merupakan bab terakhir yang berisi kesimpulan dari apa yang telah diuraikan pada bab sebelumnya serta saran untuk perkembangan penelitian yang akan datang.