

1. Pendahuluan

Pada bagian ini terdiri beberapa sub-bagian diantaranya adalah latar belakang, topik dan batasannya, tujuan dan organisasi penulisan.

Latar Belakang

Pada tahun 2015, umat muslim di seluruh dunia mencapai 1.8 miliar orang. Al-Quran merupakan kitab suci dan pedoman hidup bagi umat muslim [1]. Mempelajari dan memahami ayat Al-Quran adalah kewajiban bagi umat muslim. Dengan demikian, seorang muslim wajib mempelajari kitab suci Al-Quran secara *kaffah* atau menyeluruh.

Al-Quran terdiri lebih dari 6000 ayat dan setiap ayatnya memiliki topik yang berbeda, bahkan dalam satu ayat bisa memiliki lebih dari satu topik pada setiap ayatnya [2]. Topik yang terdapat dalam Al-Quran sangatlah beragam, mulai dari sejarah islam, amal, akhlak dan yang lainnya. Dalam tafsir Al-Quran Cordova terbitan Syaamil Quran, Bandung, terdapat 15 topik yang berbeda.

Salah satu cara yang dapat dilakukan untuk mempermudah mempelajari Al-Quran adalah dengan mengelompokkan topik-topik yang ada. Oleh karena itu, perlu dilakukan proses klasifikasi pada ayat Al-Quran. Klasifikasi ayat dalam Al-Quran juga dapat dikategorikan dalam klasifikasi teks *multi label* [1]. *Multi label* bermaksud bahwa klasifikasi akan membagi setiap ayat Al-Quran ke lebih dari 1 topik. Dengan adanya system klasifikasi Al-Quran ini, umat muslim di dunia diharapkan dapat dengan mudah membedakan dan mempelajari kategori suatu ayat dengan ayat yang lainnya.

Beberapa penelitian mengenai klasifikasi *multi label* dengan dataset Al-Quran terjemahan bahasa inggris sudah pernah dilakukan sebelumnya. Beberapa penelitian dilakukan dengan algoritma naïve bayes dan juga multi layer perceptron. Model klasifikasi dengan menggunakan algoritma naïve bayes menghasilkan performa yang cukup tinggi [2]–[5]. Model klasifikasi dengan menggunakan metode *ensemble* atau gabungan pun terbukti dapat menghasilkan hasil yang lebih baik [10] dan [11].

Pada penelitian tugas akhir ini, metode *ensemble* dipilih dengan harapan dapat meningkatkan performansi dari algoritma naïve bayes itu sendiri. Sedangkan algoritma naïve bayes dipilih dikarenakan pada penelitian lain dapat menghasilkan performansi yang cukup baik [3] dan algoritma naïve bayes juga memiliki perhitungan peluang yang sederhana sehingga memiliki waktu komputasi yang singkat. Oleh karena itu peneliti menggunakan metode *ensemble* yang di dalamnya dimasukan beberapa model dengan algoritma naïve bayes.

Topik dan Batasannya

Sebagai sumber hukum utama, topik-topik Al-Quran pada yang seharusnya dapat dipelajari oleh seluruh umat muslim di dunia. Pada kasus tersebut, peneliti membangun mudel klasifikasi multi-label pada aya Al-Quran dalam terjemahan Bahasa Inggris. Dataset yang digunakan adalah dataset al-Quran yang sudah terlabeli dengan 15 topik berdasarkan tafsir Al-Quran Cordova terbitan Syaamil Quran, Bandung.

Penelitian ini melewati beberapa tahap praproses data yang diantaranya adalah *case folding*, *punctual removal*, *stemming*, *stopword removal*, dan juga tokenisasi. Pada proses fitur ekstraksi dilakukan perhitungan TF-IDF. Algoritma yang digunakan pada penelitian ini hanya sebatas 4 algoritma naïve bayes, diantaranya adalah Gaussian naïve bayes, Multinomial naïve bayes, Complement naïve bayes dan Bernoulli naïve bayes. Beberapa algoritma naïve bayes tersebut dimasukan pada metode *ensemble*. Selain itu, hal terakhir yang menjadi batasan adalah penggunaan *hamming loss* pada tahap evaluasi.

Tujuan

Penelitian tugas akhir ini dilakukan dengan tujuan untuk membangun sebuah sistem model klasifikasi dengan metode *ensemble* dan beberapa algoritma naïve bayes. Model tersebut bertujuan mengklasifikasikan Al-Quran terjemahan Bahasa Inggris ke dalam topik-topik tertentu.

Penelitian ini dilakukan dengan tujuan untuk menganalisa terkait pengaruh penggunaan beberapa kombinasi tahap praproses terhadap hasil performansi *hamming loss*. Penelitian ini juga bertujuan untuk menganalisa pengaruh algoritma naïve bayes dengan metode *ensemble* dan tanpa metode *ensemble*.

Organisasi Tulisan

Pada penulisan pendahuluan jurnal tugas akhir ini akan dilanjutkan mengenai studi terkait yang di dalamnya menjelaskan terkait penelitian-penelitian sebelumnya dengan dataset maupun tugas serupa. Pada bagian kedua ini juga dijelaskan mengenai hal-hal yang masih berkaitan terhadap batasan pekerjaan yang dilakukan. Pada bagian ketiga akan dijelaskan mengenai metode-metode yang digunakan dan sistem yang dibangun seperti algoritma naïve bayes, metode *ensemble* dan juga tahap praproses secara lebih mendalam. pada proses klasifikasi akan dijelaskan. Pada bagian keempat yaitu evaluasi akan ditampilkan dan dijelaskan hasil pengujian serta analisis hasil pengujian dari penelitian yang telah dilakukan. Pada bagian yang terakhir yaitu bagian 5 akan dijelaskan mengenai kesimpulan dari penelitian ini dan saran untuk penelitian selanjutnya.