

KLASIFIKASI MULTI LABEL AYAT AL-QURAN MENGUNAKAN MAJORITY VOTING

Bagus Hariyadi¹, Kemas Muslim Lhaksana²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹bagushariyadi@students.telkomuniversity.ac.id, ²pembimbing1@telkomuniversity.ac.id,

Abstrak

Kitab suci Al-Quran adalah panduan dan pedoman bagi umat muslim. Al-Quran membahas banyak hal di dalamnya seperti arkanul islam, amal, sains, dan lain-lain. Dengan demikian, untuk mempermudah proses pemahaman ayat suci Al-Quran diperlukan klasifikasi topik dari masing-masing ayat. Model yang dibuat dalam penelitian ini adalah model klasifikasi multi-label yang dibangun dengan metode majority voting. Metode ini dipilih karena secara umum memberikan performa model yang lebih baik dibandingkan hanya menggunakan satu model dasar saja. Setelah itu, performa model akan diukur menggunakan dua jenis ukuran yaitu hamming loss dan mikro F1-score. Berdasarkan hasil pengujian dan eksperimen yang dilakukan performa terbaik dari majority voting diraih ketika menggunakan kombinasi model Random Forest, SVM, dan Adaptive Boosting. Kombinasi model tersebut meraih hamming loss sebesar 0.0878 dan mikro F1-Score sebesar 0.5143. Selain itu, metode majority voting dibuktikan dapat meningkatkan performa dibandingkan dengan penggunaan model secara individu.

Kata kunci : Al-Quran, topik, majority voting

Abstract

The holy book of Al-Quran is a guide and manual for Muslims. Al-Quran covers various topics such as the iman, amal, science, and more. Therefore, to facilitate the understanding of the verses of Al-Quran, it is important to categorize the topics of each verse. The model created in this study is a multi-label classification model developed using the majority voting method. This method was chosen because it generally provides better model performance than just one basic model. Subsequently, the model's performance is measured using two types of measures hamming loss and micro F1-score. Based on the experiments, the best performance is achieved using a combination of Random Forest, SVM, and Adaptive Boosting models. This combination achieved 0.5143 for the micro F1-Score and 0.0878 for the hamming loss. Additionally, the majority voting method is proven to improve performance compared to the models used individually.

Keywords: Al-Quran, topic, majority voting

1. Pendahuluan

Latar Belakang

Kitab suci Al-Quran merupakan kitab yang diturunkan kepada Nabi Muhammad SAW. Salah satu keistimewaan Al-Quran adalah mukjizat yang akan tetap relevan selamanya. Hal ini dapat dilihat pada aspek bahasa arab yang sempurna, tidak mungkin untuk ditiru, dan bukti nyata Al-Quran dalam penemuan ilmiah pada saat ini [1]. Al-Quran terdiri dari 114 surah dan 6236 ayat [2]. Ayat-ayat tersebut bisa berisi ajakan berbuat kebaikan, larangan, aturan bermasyarakat, dan lain-lain. Terlebih lagi, satu ayat dapat memiliki lebih dari satu kategori sehingga klasifikasi yang harus dilakukan adalah klasifikasi multi-label [3].

Secara umum untuk menyelesaikan masalah tersebut perlu digunakan supervised learning jenis klasifikasi perlu digunakan agar dapat melakukan klasifikasi secara otomatis. Akan tetapi, penggunaan supervised learning secara tradisional tidak mampu untuk mengkategorikan data yang bersifat multi-label seperti data ayat Al-Quran [4]. Melihat hal tersebut perlu digunakan paradigma multi-label learning untuk menjadi solusi atas masalah klasifikasi multi-label. Perbedaan pendekatan supervised learning dan multi-label learning terletak pada data yang digunakan untuk pelatihannya. Supervised learning memiliki satu kategori untuk satu observasi, sedangkan multi-label learning satu observasi dapat memiliki lebih dari satu kategori [5]. Setelah mengetahui pendekatan yang akan digunakan langkah selanjutnya adalah melakukan pemodelan menggunakan model pembelajaran mesin.

Model pembelajaran mesin yang digunakan dalam penelitian ini adalah model berjenis ensemble learning. Ensemble learning adalah metode pemodelan yang menggabungkan dua atau lebih model untuk meraih performa yang lebih baik daripada performa model ketika digunakan secara individu [6]. Metode ensemble yang digunakan dalam penelitian ini adalah Majority Voting dengan model dasar Random Forest, SVM, dan logistic regression. Model dasar tersebut digunakan karena popularitas dan performa yang baik dalam melakukan klasifikasi teks [7].