
Klasifikasi *Multilabel* pada Topik Ayat Al-Qur'an Menggunakan *Random Forest* dan *Naïve Bayes*

Imran Zulkarnaen¹, Kemas Muslim Lhaksana²

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

¹imranzulkarnaen@students.telkomuniversity.ac.id,

²kemasmuslim@telkomuniversity.ac.id

1. Pendahuluan

Al-Qur'an, sebagai kitab suci umat Islam, memiliki kedalaman makna yang mencakup berbagai aspek kehidupan, meliputi akidah, ibadah, serta etika sosial [1]. Pemahaman terhadap ayat-ayat Al-Qur'an secara tematik menjadi sangat penting untuk membantu individu dan masyarakat dalam memahami pesan-pesan yang terkandung di dalamnya secara lebih terstruktur. Dalam era digital yang semakin maju, teknologi informasi menyediakan berbagai metode untuk mendukung studi Al-Qur'an, salah satunya melalui teknik pengklasifikasian teks.

Klasifikasi teks merupakan sebuah bidang ilmu yang berkembang pesat dalam era data besar (*big data*) [2]. Metode ini memungkinkan pengelompokan informasi secara otomatis berdasarkan pola-pola tertentu, yang selanjutnya mempermudah analisis terhadap volume data yang besar. Namun, klasifikasi teks dalam konteks keagamaan, seperti pada ayat-ayat Al-Qur'an, menghadapi tantangan yang unik [3]. Bahasa Al-Qur'an yang kompleks, kaya akan makna kontekstual dan hubungan semantik, memerlukan pendekatan yang lebih canggih daripada sekadar analisis statistik tradisional.

Perkembangan metode berbasis kecerdasan buatan, seperti *machine learning* dan *graph mining*, membuka peluang baru dalam pengelolaan teks Al-Qur'an. Dengan memanfaatkan representasi graf untuk menganalisis hubungan antar kata dan algoritma prediksi yang canggih, diharapkan proses klasifikasi ayat-ayat Al-Qur'an dapat dilakukan dengan tingkat akurasi yang lebih tinggi [4]. Penelitian ini berupaya mengeksplorasi metode tersebut, tidak hanya sebagai kontribusi akademik, tetapi juga sebagai solusi praktis dalam mendukung studi tematik Al-Qur'an di era modern.

Latar Belakang

Klasifikasi ayat-ayat Al-Qur'an ke dalam kategori tematik merupakan salah satu permasalahan utama dalam bidang analisis teks keagamaan. Ayat-ayat Al-Qur'an memiliki karakteristik bahasa yang kompleks, kaya akan makna kontekstual, serta mengandung hubungan semantik yang erat antara kata-kata di dalamnya [5]. Dalam upaya untuk meningkatkan pemahaman terhadap Al-Qur'an secara lebih terstruktur, klasifikasi ini dapat memberikan manfaat yang signifikan, terutama dalam mempermudah studi tematik bagi para peneliti, akademisi, dan pembaca umum.

Pendekatan yang diadopsi dalam penelitian ini, yaitu penambangan graf (*graph mining*), menawarkan metode inovatif untuk menganalisis relasi antar kata dalam bentuk graf. Graf tersebut memfasilitasi pemetaan hubungan semantik melalui metrik sentralitas, seperti *degree*, *betweenness*, dan *closeness*, yang berperan

penting dalam identifikasi kata kunci dan struktur yang signifikan dalam teks. Berdasarkan penelitian sebelumnya, penerapan teknik penambangan graf dalam analisis teks dapat memperdalam pemahaman terhadap struktur semantik yang kompleks [6]. Teknik ini menjadi menarik karena memiliki kemampuan untuk merepresentasikan hubungan antar kata secara visual dan analitis, sehingga memberikan perspektif baru dalam memahami teks suci yang kompleks, seperti Al-Qur'an. Selain itu, penelitian menunjukkan bahwa visualisasi tematik berbasis *knowledge graph* dapat memberikan kontribusi dalam pengambilan sumber daya ilmiah dan analisis yang lebih efisien [7].

Penggunaan algoritma *Random Forest* menawarkan fleksibilitas dan stabilitas dalam klasifikasi data berlabel ganda, dengan hasil yang menunjukkan akurasi mencapai 90,77% dalam analisis sentimen [8]. Sementara itu, *Naïve Bayes* merupakan algoritma yang sederhana namun efektif untuk pengolahan teks, yang dapat berfungsi sebagai bahan perbandingan. Penelitian sebelumnya juga menunjukkan bahwa *Naïve Bayes* sering kali lebih unggul dalam klasifikasi data teks jika dibandingkan dengan algoritma lain, seperti *K-Nearest Neighbor* [9]. Evaluasi kinerja ketiga algoritma menggunakan metrik *Hamming loss* dirancang untuk memberikan wawasan yang komprehensif tentang efektivitas pendekatan yang diusulkan. Hasil perbandingan ini dapat memberikan informasi penting terkait pemilihan algoritma yang tepat untuk tugas klasifikasi tertentu.

Topik dan Batasannya

Pengklasifikasian ayat-ayat Al-Qur'an ke dalam kategori tematik merupakan suatu tantangan yang diakibatkan oleh kompleksitas bahasa serta hubungan antar kata. Dalam hal ini, *graph mining* dimanfaatkan untuk merepresentasikan hubungan antara kata-kata dalam bentuk graf, dengan sentralitas graf sebagai fitur utama.

Penelitian ini berfokus pada pengklasifikasian ayat-ayat Al-Qur'an ke dalam delapan kategori tematik dengan memanfaatkan teknik *graph mining* untuk menganalisis hubungan antar kata. Penelitian ini menggunakan algoritma *Random Forest* dan *Naïve Bayes* untuk prediksi multilabel. Isu lain yang diangkat adalah bagaimana penghapusan kata-kata umum (*stop words*) berdampak pada pengukuran sentralitas graf dan klasifikasi, serta jenis sentralitas dan metode mana yang paling efektif dalam mencapai prediksi yang akurat berdasarkan metrik *Hamming loss*.

Tujuan

Penelitian ini bertujuan untuk mengelompokkan ayat-ayat Al-Qur'an ke dalam delapan kategori tematik dengan memanfaatkan teknik *graph mining* serta algoritma *Random Forest* dan *Naïve Bayes*. Penelitian ini akan fokus pada evaluasi dampak *stop words* terhadap pengukuran sentralitas graf dan mengidentifikasi jenis sentralitas yang paling efektif dalam ekstraksi fitur. Selain itu, penelitian ini juga berusaha untuk membandingkan kinerja kedua algoritma klasifikasi tersebut dengan menggunakan metrik *Hamming loss* dan menentukan metode mana yang memberikan hasil terbaik dalam klasifikasi topik ayat-ayat Al-Qur'an.

Organisasi Tulisan

Setelah bagian Pendahuluan, penelitian ini disusun dalam beberapa komponen. Komponen pertama adalah studi terkait, yang mengulas penelitian-penelitian sebelumnya yang berkaitan dengan klasifikasi teks, khususnya dalam konteks pengklasifikasian ayat-ayat Al-Qur'an, serta berbagai teknik yang digunakan, seperti penambangan graf (*graph mining*), pembelajaran mesin (*machine learning*), dan analisis semantik. Selanjutnya, pada bagian menjelaskan sistem yang dibangun, disajikan secara rinci mengenai pendekatan yang digunakan, yaitu penambangan graf untuk menganalisis hubungan antar kata dalam ayat-ayat Al-Qur'an, serta penerapan algoritma *Random Forest* dan *Naïve Bayes* untuk klasifikasi multilabel. Dalam bagian Evaluasi, hasil-hasil eksperimen dianalisis dan dibandingkan, dengan penekanan pada dampak penghapusan kata henti (*stop words*) terhadap pengukuran sentralitas graf serta perbandingan kinerja kedua algoritma dalam klasifikasi tematik ayat-ayat Al-Qur'an. Bagian tentang Eksperimen dan Hasil menyajikan eksperimen yang dilakukan, termasuk pengolahan data dan analisis hasil yang diperoleh dengan menggunakan metrik *Hamming loss*. Akhirnya, bagian Kesimpulan dan Saran merangkum temuan dari penelitian ini dan memberikan rekomendasi untuk penelitian lebih lanjut serta penerapan praktis dari hasil yang diperoleh.