

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Hadis sebagai salah satu sumber hukum Islam setelah Al-Qur'an, memiliki peran yang sangat penting dalam kehidupan umat Islam sehari-hari. Hadis merupakan segala sesuatu yang melekat pada Rasulullah, baik berupa ucapan, perbuatan, maupun persetujuan beliau terhadap suatu tindakan atau kejadian tertentu [1]. Tidak hanya berisi ajaran dan perintah *Rasulullah*, tetapi juga mencakup berbagai aspek kehidupan, mulai dari akidah, fikih, kisah, adab, dan berbagai aspek lainnya yang menjadi pedoman bagi umat Islam dalam menjalankan kehidupan sehari-hari. Keberadaan hadis sangat krusial dalam memahami ajaran Islam secara lebih komprehensif, terutama dalam hal penjelasan terhadap ayat-ayat Al-Qur'an, karena banyak hukum Islam yang bersumber dari penafsiran hadis.

Setiap hadis terdiri atas dua komponen utama yang menjadi dasar dalam validitas dan keautentikannya. Komponen pertama adalah *Matn*, yaitu teks utama dari sebuah hadis yang merupakan isi atau pesan yang ingin disampaikan. *Matn* ini bisa berupa perintah, larangan, anjuran, atau penjelasan tentang suatu hukum dalam Islam. Komponen kedua adalah *sanad*, yaitu jalur periwayatan hadis yang menghubungkan matn hadis dengan sumber aslinya, yaitu *Rasulullah*. Sanad mencatat nama-nama perawi atau perantara yang meriwayatkan hadis dari satu generasi ke generasi berikutnya hingga sampai kepada imam hadis yang kemudian mendokumentasikan hadis tersebut dalam kitabnya [2]. Arsitektur hadis yang mencakup kedua komponen ini sebagaimana ditunjukkan pada Tabel 1.1.

Tabel 1.1 . Arsitektur hadis

حَدَّثَنَا يَحْيَى بْنُ سَعْدٍ، عَنْ أَبِيهِ، عَنِ الْقَاسِمِ بْنِ مُحَمَّدٍ، عَنْ عَائِشَةَ رَضِيَ اللَّهُ عَنْهَا قَالَتْ: قَالَ رَسُولُ اللَّهِ ﷺ: مَنْ أَخَذَ فِي أَمْرِنَا هَذَا مَا لَيْسَ فِيهِ فَهُوَ رَدٌّ (صحيح البخاري - ٢٦٩٧)

Pada tabel di atas, *matn* hadis ditunjukkan dengan teks yang dicetak tebal, sedangkan *sanad* ditunjukkan dengan teks yang dicetak miring.

Seiring dengan berkembangnya teknologi dan meningkatnya jumlah koleksi hadis yang tersedia, baik dalam bentuk teks cetak maupun digital, muncul tantangan besar

dalam mengorganisir, mengelompokkan, dan menganalisis hadis-hadis tersebut secara efisien. Salah satu tantangan utama adalah bagaimana mengklasifikasikan hadis berdasarkan tema atau topiknya secara otomatis. Hal ini menjadi semakin kompleks mengingat satu hadis dapat memiliki lebih dari satu tema atau topik yang saling terkait. Misalnya, hadis yang berbicara tentang ibadah shalat tidak hanya berkaitan dengan fikih, tetapi juga dengan aspek akidah serta nilai-nilai adab dalam beribadah maupun informasi serta kisah.

Telah ada beberapa penelitian yang melakukan klasifikasi multilabel terhadap hadis, terutama pada teks terjemahan dalam Bahasa Indonesia. Penelitian-penelitian tersebut umumnya mengkategorikan hadis ke dalam topik Anjuran, Larangan, dan Informasi dengan menggunakan berbagai metode baik *machine learning* maupun *deep learning*. Penelitian yang akan dilakukan mencoba mengembangkan klasifikasi multilabel dengan mengelompokkan hadis ke dalam topik-topik yang lebih spesifik dan terkait langsung dengan terminologi keislaman, seperti Akidah, Fikih, Kisah dan Adab. Pendekatan klasifikasi seperti ini diharapkan dapat memberikan gambaran yang lebih akurat terhadap kandungan hadis berdasarkan konsep-konsep dasar dalam Islam, sehingga mempermudah pemanfaatan hadis dalam berbagai konteks akademik maupun praktis.

Kemajuan dalam bidang *Natural Language Processing* (NLP) dalam beberapa tahun terakhir telah membuka peluang baru untuk menganalisis teks, termasuk teks-teks keislaman seperti hadis. Berbagai algoritma modern dalam NLP telah diterapkan untuk tugas klasifikasi teks, di antaranya *Recurrent Neural Network* (RNN) dan *Transformer Neural Network*. RNN memiliki kemampuan untuk mempertahankan informasi dari urutan input sebelumnya, sehingga sangat cocok untuk analisis teks yang memiliki keterkaitan makna dalam urutannya. Sementara itu, Transformer, dengan arsitektur yang lebih kompleks dan pendekatan berbasis mekanisme perhatian (*self-attention*), mampu memahami hubungan antar kata dalam teks dengan lebih baik dan efisien. Kedua pendekatan ini telah menunjukkan hasil yang menjanjikan dalam berbagai tugas NLP, termasuk klasifikasi multi label. Selain tantangan dalam metode klasifikasi, aspek bahasa juga menjadi faktor penting dalam analisis hadis. Bahasa Arab, sebagai bahasa asli dari teks-teks

keislaman termasuk Al-Qur'an dan hadis, memiliki struktur gramatikal dan morfologi yang kompleks. Bahasa Arab memiliki karakteristik seperti bentuk kata yang dapat berubah tergantung pada konteksnya (*morphological richness*), sistem akar kata yang unik (*root-based morphology*), serta adanya makna yang bisa berbeda tergantung pada harakat yang digunakan serta konteks dari kalimatnya. Hal ini membuat pemrosesan teks dalam bahasa Arab menjadi lebih menantang dibandingkan dengan bahasa lain. Oleh karena itu, dalam penerapan NLP pada hadis, diperlukan teknik khusus untuk menormalisasi kata-kata dalam bahasa Arab agar lebih mudah diproses oleh model *machine learning* tanpa kehilangan makna serta konteksnya.

Seiring dengan berkembangnya teknologi, terutama dalam bidang NLP dan meningkatnya jumlah dataset hadis digital, penelitian dalam klasifikasi hadis berbasis *deep learning* menjadi sebuah kebutuhan. Penerapan metode seperti RNN dan Transformer dalam klasifikasi multi label dapat membantu dalam mengelompokkan hadis secara lebih efisien dan akurat, sehingga mempermudah akses dan pemahaman terhadap berbagai tema yang terkandung dalam hadis.

1.2. Rumusan Masalah

Dalam penelitian ini, permasalahan yang akan dibahas adalah pengembangan model klasifikasi multi label topik hadis menggunakan algoritma Recurrent Neural Network (RNN) dan Transformer Neural Network. Pengklasifikasian topik dalam hadis merupakan tugas yang kompleks karena satu teks hadis sering kali mengandung lebih dari satu topik. Dengan demikian, dibutuhkan pendekatan yang mampu mengidentifikasi dan mengelompokkan teks hadis secara akurat ke dalam beberapa kategori yang relevan.

1.3. Tujuan dan Manfaat

Pada penelitian ini akan diimplementasikan metode-metode klasifikasi RNN dan *Transformer* dalam melakukan klasifikasi teks multilabel terhadap topik hadis dengan menggunakan dataset yang telah dikumpulkan dari ensiklopedi [3], kemudian melakukan analisis terhadap akurasi setiap model guna mengetahui metode mana yang lebih efektif dalam mendeteksi pola kalimat dari setiap hadis serta mengklasifikasikannya ke dalam beberapa topik. Selain itu, penelitian ini juga

bertujuan untuk mengkaji bagaimana kedua metode tersebut menangani ketidakseimbangan label dalam dataset, di mana terdapat perbedaan jumlah data pada setiap kategori topik. Ketidakseimbangan ini dapat menyebabkan bias dalam prediksi model.

Hadis-hadis yang akan digunakan dalam penelitian ini hanya hadis yang memiliki topik berkaitan dengan akidah, fikih, kisah dan adab. Rincian dataset tersebut seperti ditunjukkan pada Tabel 1.2.

Tabel 1.2 Rincian dataset

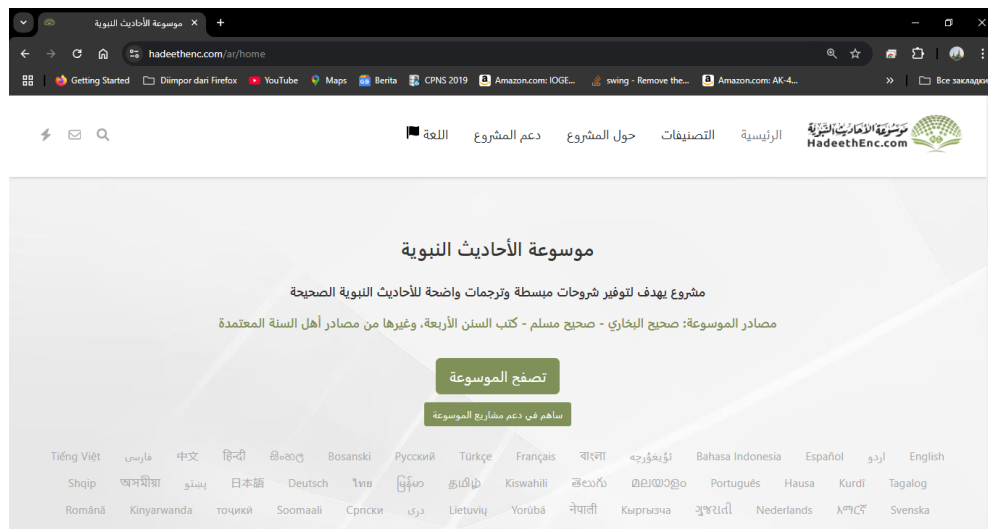
Jumlah seluruh hadis	2.223
Jumlah hadis dengan label akidah	502
Jumlah hadis dengan label fikih	1.465
Jumlah hadis dengan label kisah	1.405
Jumlah hadis dengan label adab	513

1.4. Batasan Masalah

Penelitian ini berfokus pada klasifikasi multi label hadis menggunakan algoritma RNN dan *Transformer*. Batasan yang ditetapkan dalam penelitian adalah bahwa analisis teks hadis hanya dilakukan terhadap kumpulan hadis yang terdapat dalam kitab-kitab hadis yang telah terdokumentasi secara resmi. Dataset yang digunakan dalam penelitian ini dikumpulkan dari laman <https://hadeethenc.com/>, sebuah situs yang menyediakan kumpulan hadis-hadis *Rasulullah* serta terjemahannya dalam berbagai bahasa.

Situs HadeethEnc merupakan bagian dari proyek ensiklopedi Hadis-Hadis Nabawi, yang bertujuan untuk menyediakan akses yang lebih luas terhadap hadis-hadis yang telah dikompilasi, dikategorikan dan diterjemahkan ke dalam berbagai bahasa dunia. Ensiklopedi ini memudahkan para peneliti dan masyarakat umum dalam mempelajari hadis secara lebih sistematis dan terstruktur. Proyek ini didukung serta dioperasikan oleh berbagai organisasi kajian keislaman dan lembaga wakaf yang berbasis di Kerajaan Saudi Arabia. Organisasi-organisasi ini berperan dalam memastikan keakuratan hadis yang dikumpulkan, baik dari segi teks maupun

terjemahannya, serta menyediakan referensi yang sahih bagi umat Islam di seluruh dunia [3].



Gambar 1 Homepage <https://hadeethenc.com/> [3]

Hadis yang dikumpulkan dari ensiklopedi tersebut sebelumnya telah dilakukan klasifikasi berdasarkan topik-topik tertentu oleh ahli pada bidang hadis. Diantara klasifikasi-klasifikasi tersebut antara lain seperti tercantum pada tabel Tabel 1.3.

Tabel 1.3 Daftar Klasifikasi Hadis

No	Kalsifikasi Hadis
1.	العقيدة (Akidah)
2.	الفقه وأصوله (Fikih dan <i>ushul</i> fikih)
3.	السيرة والتاريخ (Kisah dan sejarah)
4.	القرآن الكريم وعلومه (Al-Qur'an dan ilmunya)
5.	الحديث وعلومه (Hadis dan ilmu hadis)
6.	الدعوة والحسبة (Dakwah dan <i>Hisbah</i>)
7.	الفضائل والآداب (Keutamaan dan Adab)

Penelitian yang akan dilakukan akan membahas bagaimana kedua algoritma tersebut dapat digunakan untuk mengidentifikasi dan mengelompokkan hadis dengan berfokus pada empat topik utama, yaitu akidah, fikih, kisah dan adab

kemudian membandingkan tingkat akurasi antara keduanya. Selain itu, penelitian ini juga akan membahas tantangan yang dihadapi dalam proses klasifikasi tersebut, seperti ketidakseimbangan data serta kompleksitas bahasa Arab yang merupakan sebagai bahasa utama dalam literatur-literatur keislaman.

1.5. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimen dalam bidang NLP untuk melakukan klasifikasi multi-label pada teks Hadis. Algoritma yang diterapkan adalah berbagai varian dari *Recurrent Neural Network* (RNN) dan *Transformer Neural Network* untuk membandingkan performa keduanya dalam klasifikasi Hadis ke dalam beberapa kategori. Dataset yang digunakan adalah data hadis yang telah dikumpulkan dari <https://hadeethenc.com/> serta telah diberikan Akidah, Fikih, Kisah dan Adab. Prapemrosesan yang dilakukan terhadap dataset adalah diantaranya melakukan *cleaning* dan *normalization* terhadap teks hadis, yaitu dengan menghapus karakter-karakter non-Arab, mengganti abjad-abjad tertentu menjadi bentuk standarnya, serta menghapus *harakat* atau *vocal diacritic*. Kemudian dilanjutkan dengan *stopword removing* menghapus kata-kata umum yang tidak memiliki pengaruh signifikan secara makna dalam proses klasifikasi yang kemudian dilanjutkan dengan proses tokenisasi, yaitu mengubah teks hadis menjadi token-token numerik. Setelah semua proses preprosesan selesai dilakukan, dilanjutkan dengan melatih model menggunakan kedua model algoritma.

Implementasi penelitian dilakukan menggunakan bahasa pemrograman *Python* pada Google Colab dengan menggunakan framework TensorFlow, Keras, dan Transformers serta *library* tambahan seperti Pandas, NumPy, Matplotlib, dan Scikit-learn digunakan untuk analisis data dan evaluasi model. Setelah eksperimen selesai, hasil dibandingkan untuk menentukan model yang lebih unggul dalam klasifikasi Hadis berdasarkan topik yang telah tentukan sebelumnya.

1.6. Jadwal Pelaksanaan

Rencana jadwal kegiatan berdasarkan rencana kegiatan yang telah dibuat adalah sebagaimana ditunjukkan pada tabel 1.4.

Tabel 1.4. Rencana Jadwal Pelaksanaan Tugas Akhir.

Kegiatan	Bulan					
	Jul 2024	Ags 2024	Sep 2024	Okt 2024	Nov 2024	Des 2024
Persiapan dan Studi Literatur						
Perancangan Model						
Implementasi Model						
Eksperimen dan Pengujian						
Analisis Hasil dan Penyempurnaan Model						
Penulisan Laporan						