

UCAPAN TERIMA KASIH

Puji syukur kehadirat Allah yang Maha Kuasa atas segala rahmat, hidayah, dan pertolongan-Nya sehingga saya dapat menyelesaikan laporan tugas akhir ini dengan baik. Dalam proses penyusunan laporan tugas akhir ini, saya mendapatkan banyak dukungan, bimbingan, serta bantuan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, saya ingin menyampaikan penghargaan yang setinggi-tingginya kepada:

1. Orang tua dan keluarga saya yang selalu memberikan doa, dukungan, kasih sayang, serta motivasi yang tiada henti.
2. Dosen pembimbing saya, Bapak Dr. Ir. Moch. Arif Bijaksana, M. Tech., dan Bapak Bunyamin, S.Kom., M.Kom., yang dengan sabar memberikan arahan, bimbingan, serta masukan yang sangat berharga selama penelitian dan penyusunan tugas akhir ini.
3. Para dosen fakultas informatika Universitas Telkom, yang telah memberikan ilmu dan wawasan selama masa perkuliahan.
4. Teman-teman seperjuangan, khususnya Mas Yogi dan Mas Irhaz, yang selalu memberikan semangat, bantuan, serta kebersamaan yang berarti selama perkuliahan dan penyusunan tugas akhir ini.

Seluruh pihak yang telah membantu, baik secara langsung maupun tidak langsung, yang tidak dapat saya sebutkan satu per satu.

DAFTAR ISI

LEMBAR PENGESAHAN	ii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR.....	vi
UCAPAN TERIMA KASIH	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan dan Manfaat	3
1.4. Batasan Masalah	4
1.5. Metode Penelitian	5
1.6. Jadwal Pelaksanaan	6
BAB 2 TINJAUAN PUSTAKA	7
2.1. Hadis.....	7
2.2. Klasifikasi Teks.....	7
2.3. <i>Recurrent Neural Network</i>	7
2.4. <i>Long Short Term Memory (LSTM)</i>	8
2.4.1. <i>Forget Gate</i>	9
2.4.2. <i>Forget Gate</i>	9
2.4.3. <i>Output Gate</i>	9
2.4.4. <i>Memory Cell</i>	10
2.5. <i>Gated Recurrent Unit</i>	10
2.6. <i>Transformer Neural Network</i>	10
BAB 3 PERANCANGAN SISTEM	13
3.1. Desain Perancangan Sistem.....	13
3.2. Diagram Alur	13
3.3. Proses Klasifikasi	14

BAB 4	HASIL PERCOBAAN DAN ANALISIS.....	18
4.1.	Skenario Percobaan	18
4.2.	Hasil Model RNN	18
4.3.	Hasil Model <i>Transformer</i>	18
4.4.	Analisis	19
BAB 5	KESIMPULAN DAN SARAN	21
5.1.	Kesimpulan	21
5.2.	Saran	21
	DAFTAR PUSTAKA.....	23
	LAMPIRAN.....	26

DAFTAR GAMBAR

- Gambar 2.1. Logo Universitas Telkom dengan tulisan di bawah. **Error! Bookmark not defined.**
- Gambar 2.2. Contoh Penulisan footnote. **Error! Bookmark not defined.**
- Gambar 2.3. Contoh penggunaan Ibid di dalam footnote. ... **Error! Bookmark not defined.**
- Gambar 2.4. Contoh penggunaan Op. Cit pada Footnote **Error! Bookmark not defined.**

DAFTAR TABEL

Tabel 1.1. Tabel keterkaitan antara tujuan, pengujian dan kesimpulan.....	Error!
Bookmark not defined.	
Tabel 1.2. Contoh Jadwal Pelaksanaan Tugas Akhir.....	6

hadis berdasarkan tema atau topiknya secara otomatis. Hal ini menjadi semakin kompleks mengingat satu hadis dapat memiliki lebih dari satu tema atau topik yang relevan. Misalnya, hadis yang berbicara tentang ibadah shalat tidak hanya berkaitan dengan fikih, tetapi juga dengan aspek akidah serta nilai-nilai adab dalam beribadah. Oleh karena itu, metode klasifikasi multi-label menjadi pendekatan yang tepat untuk menangani permasalahan ini.

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.

Dengan adanya perkembangan NLP dan meningkatnya jumlah dataset hadis digital, penelitian dalam klasifikasi hadis berbasis deep learning menjadi semakin relevan.

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**.

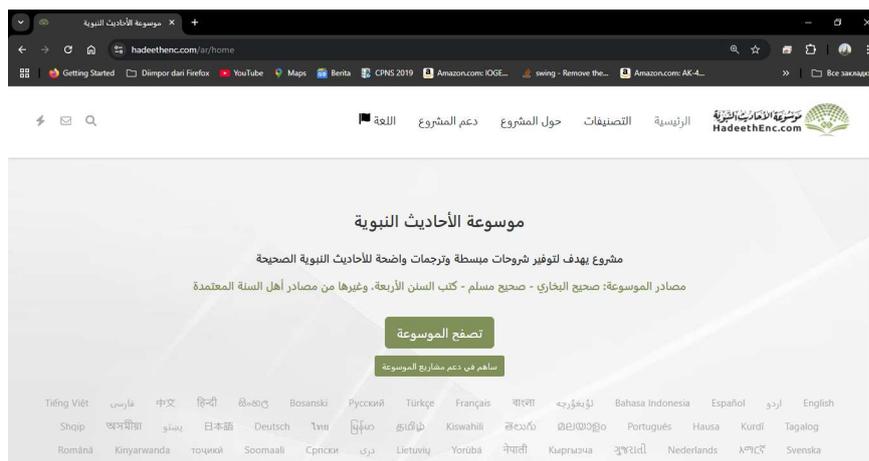
Tabel 1.2 Rincian dataset

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 preprosesn 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						

BAB 2 TINJAUAN PUSTAKA

2.1. Hadis

Secara bahasa, hadis dapat berarti sesuatu yang baru, perkataan maupun berita. Adapun secara istilah, hadis bermakna semua hal yang melekat pada Rasulullah baik berupa ucapan, perbuatan maupun persetujuan beliau. Hadis merupakan sumber hukum utama kedua dalam Islam setelah Al-Qu'an[1]. Klasifikasi hadis dapat dilakukan berdasarkan beberapa sudut pandang, diantaranya klasifikasi berdasarkan topik, klasifikasi berdasarkan tingkat kebenaran (kesahihan) maupun berdasarkan asal periwayatannya[2].

2.2. Klasifikasi Teks

Klasifikasi teks merupakan sebuah proses pelabelan terhadap suatu dokumen dengan menggunakan beberapa tahapan serta cara untuk memudahkan dalam mengatur dokumen[3]. Sedangkan menurut penelitian[4], klasifikasi teks adalah suatu pekerjaan untuk memberikan sebuah nilai boolean di setiap pasang $\langle dj, ci \rangle \in D \times C$, dimana D adalah domain dari dokumen dan $C = \{c_1, \dots, c_{|C|}\}$ adalah banyaknya kategorisasi. Nilai True akan diberikan kepada $\langle dj, ci \rangle$ apabila dokumen dj merupakan kelas ci , sedangkan nilai False akan diberikan kepada $\langle dj, ci \rangle$ apabila dokumen dj bukan merupakan kelas ci .

Adapun untuk klasifikasi teks pada Bahasa Arab memiliki tantangan unik karena memiliki kompleksitas morfologi yang cukup tinggi, yang mana sebuah kata dapat memiliki informasi signifikan yang berpengaruh pada makna kata tersebut. Oleh karena itu, diperlukan pendekatan khusus untuk menangani permasalahan tersebut, seperti yang dijelaskan dalam penelitian Arabic Text Mining. Guna memudahkan pelatihan modal, pada penelitian tersebut dilakukan normalisasi teks dengan mengubah beberapa abjad yang memiliki kesamaan bentuk atau kesamaan bunyi ke dalam bentuk dasarnya, seperti $\bar{ا}$ menjadi $ا$, $ى$ menjadi $ي$ serta $ة$ menjadi $ه$ [5].

2.3. Recurren Neural Network

Recurrent Neural Network (RNN) pertama kali diperkenalkan dari hasil penelitian yang dilakukan oleh [6] yang mana RNN merupakan sekumpulan unit input dan