**ANALISIS METODE *CONTEXTUALIZED TOPIC MODELING* (CTM) DAN BERTOPIC DALAM PEMODELAN TOPIK PADA TEKS TERJEMAHAN AL-QUR’AN BERBAHASA INDONESIA**

**Proposal Tesis**

**Diajukan untuk memenuhi syarat memperoleh gelar Magister Komputer pada Program Studi Teknik Informatika S-2**

****

OLEH:

**ASEP RIDWAN HIDAYAT**

**231012050036**

**PROGRAM STUDI TEKNIK INFORMATIKA S-2**

**PROGRAM PASCASARJANA**

**UNIVERSITAS PAMULANG**

**TANGERANG SELATAN**

**2025**

ABSTRAK

Penelitian ini bertujuan untuk menganalisis kinerja dua metode pemodelan topik modern, yaitu *Contextualized Topic Modeling* (CTM) dan BERTopic, dalam mengidentifikasi dan mengelompokkan topik pada teks terjemahan Al-Qur’an berbahasa Indonesia. Pemodelan topik merupakan teknik dalam *Natural Language Processing* (NLP) yang digunakan untuk menemukan struktur tersembunyi dalam kumpulan dokumen. Berbeda dengan pendekatan konvensional seperti Latent Dirichlet Allocation (LDA), CTM dan BERTopic memanfaatkan representasi kata berbasis konteks melalui model pembelajaran mendalam seperti BERT, sehingga mampu memahami makna kata dalam kalimat secara lebih akurat. Penelitian ini berfokus pada hasil topik yang dihasilkan serta mengevaluasi koherensi topik, relevansi konteks, dan interpretabilitas masing-masing metode dalam menangani teks religius yang kompleks dan kaya makna. Dengan menggunakan korpus terjemahan Al-Qur’an sebagai objek kajian, diharapkan dapat diperoleh wawasan mengenai kemampuan metode-metode ini dalam menyajikan representasi tematik yang bermakna secara semantik dan relevan secara teologis. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan studi Al-Qur’an berbasis teknologi NLP serta menjadi referensi bagi penelitian serupa dalam bidang ilmu komputer, linguistik, dan studi Islam. Selain itu, penelitian ini juga dapat membantu dalam penyajian informasi tematik Al-Qur’an yang lebih mudah diakses dan dipahami oleh masyarakat luas.

Kata Kunci: Contextualized Topic Modeling, BERTopic, Pemodelan Topik, Natural Language Processing,

*ABSTRACT*

*This study aims to analyze and compare topic modeling methods, namely BerTopic and Latent Dirichlet Allocation (LDA), in clustering and identifying topics from the Indonesian translation of Hadith Bukhari. Topic modeling is one of the important approaches in natural language processing that allows a deeper understanding of text content by extracting key themes. BerTopic is a method that utilizes transformer-based text representation and hierarchical clustering to generate more accurate and interpretative topics, LDA is a probabilistic model that models documents as a mixture of several topics and assumes the distribution of words on a particular topic. The data collection process is carried out through Scraping from relevant translator sources, then processed using preprocessing techniques such as Lowercasing, Cleansing, Stemming Sastrawi, and Normalization. Both algorithms, LDA and BERTopic, will be applied to identify and classify key topics. In this study, both methods are applied to the Hadith Bukhari translation dataset, and compared based on several metrics such as topic coherence, classification accuracy, and ease of interpretation. The hope of this study is that each method has different advantages and disadvantages. It can also provide insight into the best method for topic modeling of religious texts with complex language.*

*Keywords: Topic Model, BerTopic, Latent Dirichlet Allocation, Hadith Bukhari*