

ANALISIS BERT-CNN UNTUK KLASIFIKASI MULTI-LABEL DISKUSI KEAGAMAAN DAN ASOSIASI DENGAN AL-QUR'AN DAN HADITS

Candra Heru Saputra^{*1}, Suhirman²

^{1,2}Universitas Teknologi Yogyakarta, Yogyakarta

Email: ¹CandraHeru.6230211005@student.uty.ac.id, ²suhirman@uty.ac.id

*Penulis Korespondensi

(Naskah masuk: 09 Maret 2025, diterima untuk diterbitkan: 15 Desember 2025)

Abstrak

Penelitian ini menerapkan *Natural Language Processing* (NLP) dengan *Bidirectional Encoder Representations from Transformers* (BERT) yang dikombinasikan dengan *Convolutional Neural Network* (CNN) untuk klasifikasi multi-label diskusi keagamaan serta mengasosiasikannya dengan ayat Al-Qur'an dan Hadits. *Dataset* yang digunakan berasal dari diskusi keagamaan dan pertanyaan jamaah kepada ustaz yang diunggah di berbagai platform digital, seperti YouTube, Facebook, Instagram, dan situs web. Model NLP berbasis BERT digunakan untuk merepresentasikan teks secara kontekstual, sementara CNN digunakan untuk mengekstraksi fitur dan melakukan klasifikasi multi-label. Eksperimen dilakukan untuk mengeksplorasi kombinasi parameter dan pendekatan *preprocessing* teks guna meningkatkan akurasi klasifikasi. Hasil menunjukkan bahwa *tuning hyperparameter* meningkatkan F1-Score pada konfigurasi parameter ke-2 (E2) dari 0.7046 menjadi 0.7789 dan pada konfigurasi parameter ke-5 (E5) dari 0.7073 menjadi 0.7734, serta menurunkan *Hamming Loss*, yang mengindikasikan peningkatan akurasi prediksi label. *Threshold* 0.40 ditemukan sebagai nilai optimal untuk keseimbangan *Precision* dan *Recall*, yang berkontribusi terhadap peningkatan *Subset Accuracy*. Penelitian ini diharapkan dapat berkontribusi dalam pengembangan teknologi NLP berbasis bahasa Indonesia untuk klasifikasi multi-label teks keagamaan serta membuka peluang penerapan dalam aplikasi kecerdasan buatan guna meningkatkan akses informasi keagamaan secara cepat dan akurat.

Kata kunci: Klasifikasi Multi-Label, BERT, CNN, Diskusi Keagamaan, NLP, Al-Qur'an, Hadits

BERT-CNN ANALYSIS FOR MULTI-LABEL CLASSIFICATION OF RELIGIOUS DISCUSSIONS AND THEIR ASSOCIATION WITH THE QUR'AN AND HADITH

Abstract

This study applies Natural Language Processing (NLP) using Bidirectional Encoder Representations from Transformers (BERT) combined with Convolutional Neural Networks (CNN) for multi-label classification of religious discussions and their association with verses of the Qur'an and Hadith. The dataset was obtained from religious discussions and congregants' questions addressed to ustaz, collected from various digital platforms such as YouTube, Facebook, Instagram, and websites. The BERT-based NLP model was employed to represent text contextually, while CNN was used to extract features and perform multi-label classification. Experiments were conducted to explore parameter combinations and text preprocessing approaches to improve classification accuracy. The results show that hyperparameter tuning increased the F1-Score in the second parameter configuration (E2) from 0.7046 to 0.7789 and in the fifth configuration (E5) from 0.7073 to 0.7734, while reducing the Hamming Loss, indicating an improvement in label prediction accuracy. A threshold of 0.40 was found to be the optimal value for balancing Precision and Recall, contributing to an increase in Subset Accuracy. This research is expected to contribute to the development of Indonesian-language NLP technology for multi-label classification of religious texts and to open opportunities for practical applications in artificial intelligence systems to enhance rapid and accurate access to religious information.

Keywords: Multi-Label Classification, BERT, CNN, Religious Discussion, NLP, Quran, Hadith

1. PENDAHULUAN

Al-Qur'an merupakan kitab suci umat Islam dan sumber utama ajaran Islam (Choirulfikri et al., 2022;

Nouh & Alsayat, 2020), yang diyakini sebagai firman Allah SWT yang diwahyukan kepada Nabi Muhammad SAW. Al-Qur'an menyediakan panduan hidup yang komprehensif, mencakup aspek spiritual,

moral, sosial, dan hukum, yang menjadi pedoman bagi umat Islam dalam menjalani kehidupan sehari-hari (Arkok & Zeki, 2021). Selain Al-Qur'an, Hadits juga merupakan sumber utama ajaran Islam yang berisi perkataan, tindakan, dan persetujuan Nabi Muhammad SAW (Kustiawan et al., 2022). Hadits berperan sebagai penjelasan dan pelengkap bagi ayat-ayat Al-Qur'an (Syakhrani, 2023), memberikan panduan moral dan etika, menjelaskan konteks serta makna ayat-ayat, serta menawarkan contoh praktis penerapan ajaran Islam (Altammami et al., 2020). Oleh karena itu, kemampuan untuk mengakses dan memahami Al-Qur'an serta Hadits dengan cepat dan akurat menjadi kebutuhan penting bagi umat Islam dalam menjalani kehidupan sehari-hari (Akbar et al., 2024).

Seiring perkembangan teknologi, diskusi keagamaan kini meluas ke platform digital seperti YouTube, Facebook, Instagram, dan situs web. Media ini memfasilitasi jamaah untuk memperoleh bimbingan dari ustaz secara lebih fleksibel. Dalam interaksi tersebut, jamaah sering mengajukan pertanyaan tentang berbagai aspek kehidupan, sementara ustaz merujuk pada ayat Al-Qur'an atau Hadits sebagai jawaban. Namun, jumlah konten yang semakin besar menyulitkan pencarian manual terhadap referensi yang relevan. Variasi topik dan cara penyampaian memperbesar tantangan pencarian, sehingga dibutuhkan sistem otomatis yang dapat melakukan klasifikasi dan menghubungkan teks dengan rujukan keagamaan.

Masalah utama dalam pencarian referensi Al-Qur'an dan Hadits adalah keterbatasan sistem konvensional dalam menghubungkan teks diskusi keagamaan dengan rujukan yang tepat (Wasfey et al., 2022). Pertanyaan jamaah sering kali kompleks dan membutuhkan lebih dari satu referensi sekaligus. Kondisi ini menuntut pendekatan klasifikasi multi-label. Pencarian manual tidak hanya lambat tetapi juga berisiko menghasilkan rujukan yang kurang tepat, sehingga diperlukan solusi berbasis kecerdasan buatan (*Artificial Intelligence/AI*). Salah satu cabang AI yang relevan adalah *Machine Learning (ML)*, khususnya bidang *Natural Language Processing (NLP)*.

Model NLP berbasis *Bidirectional Encoder Representations from Transformers (BERT)* telah terbukti efektif dalam memahami konteks teks secara mendalam (Devlin et al., 2019; Nissa & Yulianti, 2023), dan digunakan secara luas pada tugas klasifikasi teks maupun analisis sentimen (Alyoubi et al., 2023). Di sisi lain, *Convolutional Neural Networks (CNN)*, meskipun awalnya dikembangkan untuk pengolahan citra (Torres et al., 2020), juga menunjukkan kinerja tinggi dalam klasifikasi teks (Wang et al., 2021; Zhou et al., 2022). CNN unggul dalam mengekstraksi pola lokal dari *embedding* teks, sehingga potensial untuk dikombinasikan dengan BERT dalam klasifikasi multi-label.

Sejumlah penelitian terdahulu menunjukkan potensi besar BERT dan CNN untuk klasifikasi teks, meski belum menyentuh konteks diskusi keagamaan berbahasa Indonesia. Li et al. (2023) menggunakan BERT untuk klasifikasi kebijakan industri energi di China dengan F1-Score 87,74%, sementara Nissa & Yulianti (2023) menerapkan IndoBERT untuk klasifikasi multi-label ulasan hotel dengan F1-Score 92,83%. Ahmed (2022) menguji BERT multibahasa pada teks Urdu, sedangkan Arslan & Cruz (2024) mengoptimalkan *fine-tuned* BERT untuk teks bisnis dan mencapai F1-Score 97,8%. Penelitian lain mengeksplorasi kombinasi BERT dan CNN, misalnya pada komentar media sosial (Nabiilah et al., 2024) dan dokumen akademik (Cai et al., 2020), menunjukkan hasil menjanjikan tetapi tidak terkait teks keagamaan.

Dalam konteks bahasa Indonesia, beberapa studi relevan antara lain klasifikasi berita hoaks dengan mBERT (Hutama & Suhartono, 2022), clickbait dengan IndoBERT (Ahmadi & Chowanda, 2023), status COVID-19 (Budiman et al., 2024), dan kecelakaan lalu lintas (Naufal & Girsang, 2024). Kajian ini memperlihatkan potensi BERT untuk bahasa Indonesia, tetapi belum ada penelitian yang secara khusus mengintegrasikan BERT–CNN untuk klasifikasi multi-label diskusi keagamaan dengan asosiasi langsung ke Al-Qur'an dan Hadits. Gap inilah yang menjadi fokus penelitian ini.

Studi terkait teks keagamaan sebelumnya banyak berfokus pada klasifikasi topik Hadits (Nabiilah et al., 2021), validitas sanad (Nuha & Rochmawati, 2019), hingga deteksi ujaran Islamophobia di media sosial (Aldreabi & Blackburn, 2023). Akan tetapi, penelitian yang secara khusus menghubungkan teks diskusi jamaah dengan rujukan Al-Qur'an dan Hadits melalui pendekatan NLP masih jarang ditemukan. Hal ini menegaskan adanya ruang penelitian baru, khususnya dalam konteks bahasa Indonesia yang belum banyak mendapat perhatian dalam riset NLP global.

Berdasarkan latar belakang tersebut, penelitian ini mengajukan model hibrida BERT–CNN untuk klasifikasi multi-label diskusi keagamaan berbahasa Indonesia. Permasalahan utama yang diteliti adalah: (1) bagaimana penerapan kombinasi BERT–CNN dalam klasifikasi multi-label teks keagamaan, dan (2) konfigurasi parameter apa yang optimal untuk meningkatkan akurasi model.

Kontribusi penelitian ini dapat dilihat dari tiga aspek utama. Pertama, penelitian ini mengusulkan arsitektur hibrida BERT–CNN yang memanfaatkan BERT untuk representasi teks kontekstual dan CNN untuk ekstraksi pola lokal, sehingga mampu melakukan klasifikasi multi-label diskusi keagamaan. Kedua, penelitian ini menyediakan *dataset* berbahasa Indonesia yang dianotasi secara manual dari diskusi ustaz dan jamaah di platform digital, dengan label berupa ayat Al-Qur'an (QS) dan Hadits (HB). Ketiga, penelitian ini mengevaluasi pengaruh kombinasi

hyperparameter dan *threshold* terhadap performa model, dengan tujuan mengidentifikasi konfigurasi yang paling optimal untuk meningkatkan akurasi klasifikasi multi-label berbasis teks keagamaan. Dengan kontribusi ini, penelitian diharapkan memperluas kajian NLP berbasis bahasa Indonesia dalam domain keagamaan, sekaligus menyediakan dasar bagi pengembangan aplikasi praktis seperti *chatbot* atau sistem pencarian otomatis rujukan Al-Qur'an dan Hadits.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Natural Language Processing* (NLP) berbasis *Bidirectional Encoder Representations from Transformers* (BERT) yang dikombinasikan dengan *Convolutional Neural Networks* (CNN) untuk klasifikasi multi-label diskusi keagamaan. Model ini dirancang untuk menghubungkan teks pertanyaan jamaah dengan referensi ayat Al-Qur'an dan Hadits secara otomatis. *Dataset* yang digunakan dikumpulkan dari berbagai platform digital, termasuk YouTube, Facebook, Instagram, serta situs web yang berfokus pada diskusi keagamaan. Fokus pengumpulan data adalah pada interaksi antara ustaz dan jamaah, di mana ustaz memberikan jawaban yang merujuk secara eksplisit pada ayat Al-Qur'an dan Hadits.

Proses pengumpulan data dilakukan dengan metode yang berbeda tergantung pada jenis konten. Untuk video YouTube, data dikumpulkan dengan cara menonton langsung dan mencatat informasi yang relevan secara manual. Peneliti mendengarkan diskusi, mencatat pertanyaan yang diajukan oleh jamaah, serta merangkum jawaban ustaz yang mengacu pada sumber ajaran Islam. Sementara itu, untuk data dari Facebook, Instagram, dan situs web, dilakukan pengumpulan secara manual dengan membaca postingan diskusi keagamaan, mencatat informasi penting, dan memastikan relevansi konten terhadap tujuan penelitian. Semua data yang terkumpul kemudian diproses lebih lanjut melalui analisis dan ekstraksi informasi untuk menyiapkan *dataset* pelatihan model.

Setelah data terkumpul, dilakukan pemrosesan teks dengan BERT, yang dimulai dengan tokenisasi menggunakan *WordPiece Tokenizer*. Tokenisasi ini bertujuan untuk membagi teks menjadi unit-unit kecil yang lebih mudah diproses oleh model, termasuk menangani kata-kata yang tidak dikenal dengan memecahnya menjadi sub-kata. Selanjutnya, penambahan token khusus seperti '[CLS]' di awal teks dan '[SEP]' di akhir teks dilakukan untuk membantu model memahami konteks teks secara keseluruhan. Untuk memastikan keseragaman panjang input, diterapkan *padding* dan *truncation*, di mana teks yang lebih pendek dari panjang maksimum akan ditambahkan token '[PAD]', sedangkan teks yang lebih panjang akan dipotong. Setelah proses ini, token hasil tokenisasi dikonversi menjadi indeks numerik berdasarkan kosa kata BERT dan dikodekan

dengan *positional encoding*, yang mempertahankan informasi urutan dalam teks. Representasi akhir dari teks berbentuk *embedding* vektor berdimensi 768, yang kemudian digunakan sebagai input untuk model klasifikasi berbasis CNN.

Model yang dikembangkan dalam penelitian ini merupakan integrasi antara BERT sebagai representasi teks kontekstual dan CNN 1D sebagai pengklasifikasi multi-label. Representasi semantik utama diambil dari *embedding* token [CLS] hasil keluaran BERT, kemudian diteruskan ke lapisan konvolusi CNN 1D untuk mengekstraksi pola lokal yang relevan dari teks. Proses konvolusi tersebut dapat dinyatakan secara matematis pada persamaan (1).

$$h_i = f(W \cdot x_{i:i+k-1} + b) \quad (1)$$

dengan $x_{i:i+k-1}$ merupakan *embedding* kata, W adalah bobot filter, b bias, dan f fungsi aktivasi ReLU. Hasil konvolusi kemudian diproses melalui *max pooling* untuk mereduksi dimensi sekaligus mempertahankan fitur paling representatif, sebagaimana ditunjukkan pada persamaan (2).

$$h = \max(h_1, h_2, \dots, h_n) \quad (2)$$

Vektor hasil *pooling* selanjutnya diteruskan ke *fully connected layer* dan dipetakan ke *output* multi-label menggunakan fungsi aktivasi *sigmoid*, yang diformulasikan pada persamaan (3).

$$y_j = \sigma(z_j) = \frac{1}{1+e^{-z_j}} \quad (3)$$

Label diklasifikasikan sebagai positif apabila $y_j \geq \theta$, dengan θ merupakan nilai ambang (*threshold*). Dengan alur ini, BERT berperan sebagai ekstraktor representasi kontekstual, sedangkan CNN mendeteksi pola n-gram lokal yang relevan untuk mendukung klasifikasi multi-label.

Proses pelatihan model dilakukan dengan membagi *dataset* menjadi data latih dan data uji. Data latih digunakan untuk melatih model dalam mengenali pola hubungan antara pertanyaan jamaah dan referensi Al-Qur'an serta Hadits, sedangkan data uji digunakan untuk mengukur kemampuan model dalam menggeneralisasi pola-pola tersebut pada data yang belum pernah dilihat sebelumnya. Pelatihan dilakukan dalam beberapa *epoch* dengan penerapan *early stopping*, yang memastikan model berhenti pada titik optimal untuk menghindari *overfitting*. Pada tahap awal (*Baseline Model*), model dilatih menggunakan konfigurasi awal tanpa *tuning hyperparameter*, untuk memperoleh gambaran dasar tentang performa klasifikasi. Selanjutnya, dalam tahap *tuning hyperparameter*, berbagai parameter seperti *learning rate*, jumlah *filter* CNN, ukuran *kernel*, *dropout rate*, dan *optimizer* dioptimalkan guna meningkatkan akurasi model.

Evaluasi model dilakukan menggunakan berbagai metrik, termasuk *Hamming Loss*, *Precision*,

Recall, *F1-Score*, dan *Subset Accuracy*. *Hamming Loss* digunakan untuk mengukur kesalahan dalam klasifikasi multi-label, sedangkan *Precision* dan *Recall* mengevaluasi sejauh mana model dapat mengidentifikasi referensi yang relevan secara akurat dan menyeluruh. *F1-Score* memberikan gambaran menyeluruh tentang keseimbangan antara *Precision* dan *Recall*, sementara *Subset Accuracy* mengukur ketepatan prediksi dalam keseluruhan label untuk setiap sampel. Hasil evaluasi ini digunakan untuk menilai efektivitas model dalam mengklasifikasikan teks diskusi keagamaan serta untuk mengidentifikasi area yang perlu diperbaiki.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Tabel 1. Daftar Contoh Referensi Al-Qur'an dan Hadits

Label	Referensi
HB(2008)	Yahya bin Bukair telah menceritakan kepada kami: Al-Laits menceritakan kepada kami, dari 'Uqail, dari Ibnu Syihab, beliau mengatakan: Abu Salamah mengabarkan kepadaku: Bahwa Abu Hurairah <i>radhiyallahu 'anhu</i> mengatakan: Aku mendengar Rasulullah <i>shallallahu 'alaihi wa sallam</i> bersabda mengenai bulan Ramadan, "Siapa saja yang salat tarawih dalam keadaan beriman dan mengharap pahala, maka segala dosanya yang lalu akan diampuni."
HB(2009)	Abdullah bin Yusuf telah menceritakan kepada kami: Malik mengabarkan kepada kami, dari Ibnu Syihab, dari Humaid bin 'Abdurrahman, dari Abu Hurairah <i>radhiyallahu 'anhu</i> : Bahwa Rasulullah <i>shallallahu 'alaihi wa sallam</i> bersabda, "Siapa saja yang salat tarawih di bulan Ramadan dalam keadaan beriman dan mengharap pahala, maka segala dosanya yang lalu akan diampuni." Ibnu Syihab mengatakan: Rasulullah <i>shallallahu 'alaihi wa sallam</i> wafat dalam keadaan perkara (salat tarawih tidak dikerjakan secara berjemaah) tetapi dalam keadaan itu. Kemudian tetap dalam keadaan itu di masa kekhilafahan Abu Bakr dan di awal masa kekhilafahan 'Umar <i>radhiyallahu 'anhumu</i> .
QS(2:187)	Makanlah dan minumlah sampai telah jelas bagi kalian benang putih dari benang hitam, yaitu fajar. Kemudian sempurnakanlah puasa sampai (datang) malam hari.
...	...
HB(2014)	Ali bin 'Abdullah telah menceritakan kepada kami: Sufyan menceritakan kepada kami, beliau berkata: Kami telah menghafalnya, dan beliau menghafal dari Az-Zuhri, dari Abu Salamah, dari Abu Hurairah <i>radhiyallahu 'anhu</i> , dari Nabi shallallahu 'alaihi wa sallam, beliau bersabda, "Barangsiaapa puasa Ramadhan dengan iman dan mengharap pahala, niscaya diampuni dosanya yang telah lalu. Dan barangsiapa shalat malam lailatul qadr dengan iman dan mengharap pahala, niscaya diampuni dosanya yang telah lalu." Sulaiman bin Katsir menguatkan hadits ini dari Az-Zuhri.

Berdasarkan *Dataset* dalam penelitian ini diperoleh melalui proses anotasi manual dari diskusi keagamaan, dengan merujuk pada ayat Al-Qur'an dan Hadits yang secara eksplisit disebutkan oleh ustaz. Setiap teks diberi label yang mencerminkan hubungan antara pertanyaan jamaah dan referensi keagamaan yang diberikan. Mengingat sifat diskusi yang bersifat multi-label, satu teks dapat memiliki lebih dari satu label. Sebagai contoh, pertanyaan tentang shalat tarawih dapat dikaitkan dengan beberapa Hadits yang membahas keutamaannya.

Anotasi dilakukan dengan mencatat referensi yang disebutkan langsung oleh ustaz, kemudian setiap referensi diberikan kode sistematis, seperti QS(x:y) untuk ayat Al-Qur'an dan HB(z) untuk Hadits. Pemeriksaan manual dilakukan untuk memastikan keakuratan label yang diberikan, sehingga hanya referensi yang benar-benar disebutkan dalam diskusi yang dimasukkan dalam *dataset*.

Tabel 1 menampilkan contoh ayat Al-Qur'an dan Hadits yang digunakan sebagai rujukan dalam anotasi *dataset*. Sedangkan Tabel 2 menyajikan contoh daftar pertanyaan jamaah beserta label yang diberikan berdasarkan referensi keagamaan yang disebutkan dalam diskusi. Dengan pendekatan ini, *dataset* yang dihasilkan memiliki akurasi tinggi dalam mencerminkan hubungan antara pertanyaan keagamaan dan sumber rujukannya. Selain itu, anotasi berbasis referensi langsung dari ustaz memastikan bahwa label dalam model klasifikasi multi-label memiliki dasar yang kuat dan valid dalam konteks keislaman.

Pada tahap awal, dilakukan eksperimen *baseline* untuk mengevaluasi performa awal model CNN-BERT dalam klasifikasi multi-label. Model ini mengombinasikan BERT sebagai *embedding* teks dan CNN sebagai ekstraktor fitur sebelum klasifikasi. Eksperimen bertujuan untuk menguji berbagai konfigurasi *learning rate*, *dropout rate*, *pooling strategy*, jumlah *filter* CNN, dan *kernel size* guna memahami sejauh mana model dapat menggeneralisasi tugas klasifikasi.

Tabel 2. Daftar Contoh Pertanyaan Jamaah

Pertanyaan	Label
Ustaz, apakah ada keutamaan khusus bagi orang yang melaksanakan shalat tarawih berjemaah?	HB(2008), HB(2009)
Apa yang sebaiknya dilakukan jika mendengar azan saat masih ada makanan di mulut?	QS(2:187), HB(1916), HB(1917)
Mengapa puasa bisa membantu kita meningkatkan kesabaran dan pengendalian diri?	HB(1894), HB(1904)
Ustaz, apakah ada doa khusus yang dianjurkan untuk dibaca saat malam Lailatul Qadar?	HB(1901)
...	...
Ustaz, apa sebenarnya makna dari Lailatul Qadr dan mengapa malam itu begitu istimewa?	QS(97:1-5), HB(2014)

Eksperimen *baseline* menggunakan 500 *epoch* dengan *batch size* 8, serta menerapkan *early stopping*

dengan *patience* 100 *epoch* untuk mencegah *overfitting*. Lima konfigurasi model diuji dengan kombinasi *hyperparameter* yang berbeda, sebagaimana ditampilkan dalam Tabel 3.

Tabel 3. Konfigurasi Model <i>Baseline</i> CNN-BERT						
Ex p	Learni ng Rate	Drop out	Pooling Strategy	Filter s	Ker nel Size	Thresh old
E1	0.001	0.5	CLS	128	5	0.5
E2	0.0001	0.6	Mean	256, 128	5, 3	0.4
E3	0.0000 5	0.7	Max	512	5	0.35
E4	0.0000 1	0.4	Dynamic	128	5	0.5
E5	0.0002	0.5	Hierarchical	256	5	0.45

Evaluasi dilakukan menggunakan *Hamming Loss*, *Precision*, *Recall*, *F1-Score* dan *Subset Accuracy*, dengan setiap konfigurasi diuji sebanyak lima kali (*5 trials*) untuk memastikan hasil yang stabil dan dapat diandalkan. Hasil evaluasi menunjukkan bahwa model E2 dan E5 memiliki performa terbaik, dengan *Hamming Loss* terendah dan *F1-Score* tertinggi, sebagaimana disajikan dalam Tabel 4.

Dari hasil evaluasi model *baseline*, model E2 dan E5 menunjukkan performa terbaik, dengan *F1-Score* tertinggi dan *Hamming Loss* terendah, yang menunjukkan keseimbangan yang baik antara *Precision* dan *Recall*. Model E5 unggul dalam *Precision*, menandakan kemampuannya dalam menghasilkan prediksi yang lebih akurat dengan lebih sedikit *false positives*, sementara E2 memiliki *Subset Accuracy* tertinggi, yang menunjukkan bahwa model lebih sering mengklasifikasikan keseluruhan set label dengan benar dibandingkan model lainnya.

Tabel 4. Hasil Evaluasi Model <i>Baseline</i> CNN-BERT						
Ex p	Epoch	Hamming Loss (mean)	Precision (mean)	Recall (mean)	F1-Score (mean)	Subset Accuracy (mean)
E1	180	0.0602	0.9282	0.38	0.54	0.3333
				58	1	
E2	191. 2	0.0468	0.8506	0.60	0.70	0.5697
				28	46	
E3	203. 4	0.1564	0.6183	0.30	0.33	0.1545
				78	54	
E4	391. 2	0.0498	0.8816	0.53	0.66	0.4879
				62	65	
E5	155	0.0473	0.8318	0.61	0.70	0.5667
				56	73	

Sebaliknya, model E3 memiliki performa terburuk, dengan *F1-Score* terendah dan *Hamming Loss* tertinggi, menandakan kesulitan model dalam mengenali pola klasifikasi multi-label. Model E4 memiliki performa yang cukup baik, namun membutuhkan jumlah *epoch* yang jauh lebih banyak, sehingga kurang efisien dibandingkan model lainnya.

Oleh karena itu, untuk meningkatkan akurasi dan efisiensi model, dilakukan eksperimen *tuning hyperparameter* dengan fokus pada optimasi *learning rate*, *dropout rate*, dan *batch size* guna meningkatkan akurasi klasifikasi multi-label lebih lanjut.

Pada tahap *tuning hyperparameter*, berbagai kombinasi parameter diuji untuk menemukan konfigurasi optimal. *Batch size* yang diuji terdiri dari 8 dan 16, sementara *learning rate* diuji dengan nilai 8e-5, 1e-4, dan 2e-4. Untuk *dropout rate*, variasi yang digunakan adalah 0.3, 0.4, dan 0.5, guna mencegah *overfitting*. Selain itu, model diuji dengan tiga jenis *optimizer*, yaitu Adam, AdamW, dan RMSprop, serta tiga *pooling strategy*, yaitu *Dynamic*, *Max*, dan *Mean*, untuk menentukan metode terbaik dalam ekstraksi fitur CNN.

Lebih lanjut, eksperimen juga menguji dua kombinasi *filter CNN* (256 dan 128) serta dua ukuran *kernel* (5 dan 3) untuk memahami pengaruhnya terhadap performa model. Selain itu, pengaruh *threshold* dalam klasifikasi multi-label turut dievaluasi dengan tiga nilai utama, yaitu 0.35, 0.40, dan 0.50, guna menemukan keseimbangan terbaik antara *Precision* dan *Recall*.

Pendekatan *grid search* diterapkan untuk menguji setiap kombinasi secara sistematis, dengan setiap konfigurasi diuji sebanyak lima kali (*5 trials*) untuk mendapatkan hasil yang lebih stabil. Hasil *tuning hyperparameter* menunjukkan bahwa konfigurasi terbaik mampu meningkatkan akurasi model, dengan keseimbangan yang lebih optimal antara *Precision*, *Recall*, dan *F1-Score*, serta efisiensi yang lebih baik dalam jumlah *epoch* yang diperlukan untuk mencapai konvergensi.

Eksperimen *tuning hyperparameter* dilakukan untuk mengevaluasi kombinasi optimal dari *batch size*, *learning rate*, *dropout rate*, *optimizer*, *pooling strategy*, *filter CNN*, *kernel size*, dan *threshold* guna meningkatkan akurasi serta efisiensi model klasifikasi multi-label. Berdasarkan hasil *tuning* yang dirangkum dalam Tabel 5, model E2 dengan konfigurasi *batch size* 8, *learning rate* 0.0001, *dropout rate* 0.4, *optimizer* RMSprop, *pooling strategy* *dynamic*, *filter CNN* 256 & 128, *kernel size* 5 & 3, serta *threshold* 0.35 menghasilkan performa terbaik. Model ini memiliki *Hamming Loss* terendah sebesar 0.0381 dan *F1-Score* tertinggi sebesar 0.7789, menunjukkan keseimbangan optimal antara *Precision* (0.8459) dan *Recall* (0.7220). Selain itu, model ini juga memiliki *Subset Accuracy* tertinggi sebesar 0.6545.

Model E5 dengan konfigurasi *batch size* 16, *learning rate* 0.0002, *dropout rate* 0.3, *optimizer* RMSprop, *pooling strategy* *dynamic*, *filter CNN* 256, *kernel size* 5, serta *threshold* 0.5 juga menunjukkan performa yang kompetitif. Model ini mencapai *F1-Score* sebesar 0.7549, dengan *Precision* tertinggi sebesar 0.8524 dan *Recall* sebesar 0.6780, yang menunjukkan fleksibilitas model dalam menangkap lebih banyak label yang relevan.

Tabel 5. Hasil Evaluasi Eksperimen Tuning Hyperparameter

Epoch (mean n)	Ex p	Batch Size	Learning Rate	Drop out Rate	Optimizer	Pooling	Threshold	Filter	Kernel	Hamming Loss (mean)	Precision (mean)	Recall (mean)	F1-Score (mean)	Subset Accuracy (mean)
186.4	E2	8	0.0001	0.4	RMSProp	dynamical	0.35	256, 128	5, 3	0.0380	0.8458	0.7219	0.7788	0.6545
144.6	E2	8	0.0002	0.3	AdamW	dynamical	0.4	256, 128	5, 3	0.0392	0.8260	0.7319	0.7758	0.6484
191.6	E2	8	8.00E-05	0.3	RMSProp	dynamical	0.5	256, 128	5, 3	0.0405	0.8602	0.6723	0.7547	0.6151
180.8	E2	16	0.0002	0.5	RMSProp	dynamical	0.35	256, 128	5, 3	0.0396	0.8210	0.7347	0.7747	0.6121
148	E2	16	0.0002	0.3	AdamW	dynamical	0.4	256, 128	5, 3	0.0404	0.8288	0.7120	0.7658	0.6515
176.2	E2	16	0.0002	0.3	RMSProp	dynamical	0.5	256, 128	5, 3	0.0412	0.8513	0.6737	0.7518	0.6181
150.8	E5	8	0.0002	0.3	AdamW	dynamical	0.35	256	5	0.0405	0.8069	0.7460	0.7733	0.6303
147.8	E5	8	0.0002	0.3	AdamW	dynamical	0.4	256	5	0.0397	0.8288	0.7205	0.7708	0.6606
152.8	E5	8	0.0002	0.3	AdamW	dynamical	0.5	256	5	0.0409	0.8507	0.6780	0.7544	0.6212
178.8	E5	16	0.0002	0.3	RMSProp	dynamical	0.35	256	5	0.0400	0.8239	0.7248	0.7706	0.6424
176.6	E5	16	0.0002	0.3	RMSProp	dynamical	0.4	256	5	0.0411	0.8245	0.7092	0.7621	0.6303
										0.07	0.27	0.2	0.59	0.03

Model ini juga memiliki *Subset Accuracy* sebesar 0.6242 yang menunjukkan kemampuan model dalam memprediksi keseluruhan set label secara lebih akurat dibandingkan konfigurasi lainnya. Dengan demikian, model E2 dan E5 tetap menjadi pilihan utama dalam meningkatkan akurasi klasifikasi multi-label berbasis teks keagamaan.

Eksperimen *tuning hyperparameter* juga mengevaluasi dampak penggunaan *threshold* terhadap performa model dalam klasifikasi multi-label. Seperti dirangkum dalam Tabel 6, tiga nilai *threshold* utama diuji, yaitu 0.35, 0.40, dan 0.50, untuk memahami keseimbangan antara *Precision* dan *Recall* serta dampaknya terhadap metrik evaluasi lainnya.

Hasil eksperimen menunjukkan bahwa *threshold* 0.35 menghasilkan *Recall* tertinggi (0.7219) tetapi *Precision* lebih rendah (0.8459), yang mengindikasikan model lebih fleksibel dalam menangkap label yang relevan, meskipun dengan peningkatan false positives. *Threshold* 0.40 memberikan keseimbangan terbaik, dengan *F1-Score* sebesar 0.7758, *Precision* 0.8260, dan *Recall* 0.7319, menjadikannya pilihan optimal dalam klasifikasi multi-label yang membutuhkan akurasi tinggi.

Sementara itu, *threshold* 0.50 meningkatkan *Precision* hingga 0.8602 tetapi menurunkan *Recall* menjadi 0.6723, menyebabkan *F1-Score* terendah (0.7547). Model dengan *threshold* ini lebih selektif dalam menentukan label positif, tetapi lebih rentan terhadap false negatives. Secara keseluruhan, *threshold* 0.40 dipilih sebagai nilai optimal karena

memberikan keseimbangan terbaik antara *Precision* dan *Recall*, serta menghasilkan *Subset Accuracy* yang lebih baik dibandingkan *threshold* lainnya.

Tabel 6. Perbandingan Hasil Model Berdasarkan Threshold

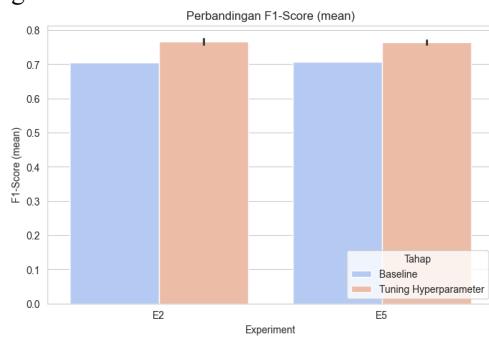
Threshold	Hamming Loss (mean)	Precision (mean)	Recall (mean)	F1-Score (mean)	Subset Accuracy (mean)
0.35	0.03807	0.84588	0.72198	0.77885	0.65454
0.4	0.03926	0.82605	0.73191	0.77583	0.64848
0.5	0.04058	0.86028	0.67234	0.75473	0.61515

Hasil *tuning hyperparameter* menunjukkan peningkatan signifikan dalam performa model, sebagaimana divisualisasikan dalam Gambar 1 hingga Gambar 5. *F1-Score* model E2 meningkat dari 0.7046 menjadi 0.7789, sedangkan model E5 meningkat dari 0.7073 menjadi 0.7734 (Gambar 1), menandakan keseimbangan yang lebih baik antara *Precision* dan *Recall*.

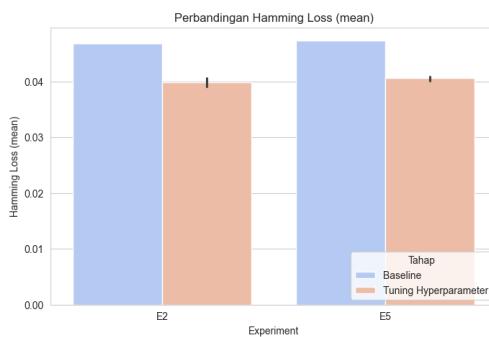
Selain itu, *tuning hyperparameter* juga berhasil menurunkan *Hamming Loss*, yang mencerminkan pengurangan kesalahan klasifikasi. Model E2 mengalami penurunan dari 0.0468 menjadi 0.0381, sedangkan E5 dari 0.0473 menjadi 0.0406 (Gambar 2). *Precision* pada model E2 sedikit menurun dari 0.8603 menjadi 0.8459, sementara E5 tetap stabil di 0.8239 (Gambar 3), menunjukkan bahwa meskipun ada sedikit kompromi pada *Precision*, peningkatan *Recall* lebih signifikan, dengan E2 naik dari 0.6723

menjadi 0.7220 dan E5 dari 0.6780 menjadi 0.7461 (Gambar 4).

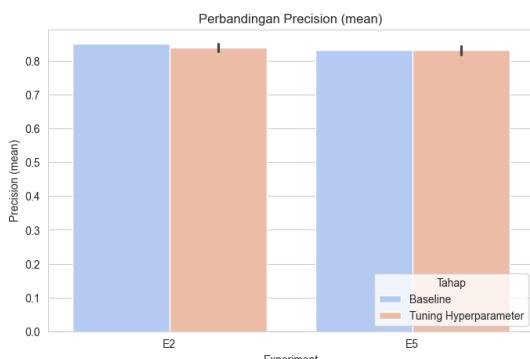
Selain itu, *Subset Accuracy* juga meningkat, dengan E2 naik dari 0.6151 menjadi 0.6545 dan E5 dari 0.6212 menjadi 0.6606 (Gambar 5), yang menunjukkan bahwa model lebih sering memberikan prediksi yang benar secara keseluruhan. Secara keseluruhan, *tuning hyperparameter* meningkatkan *F1-Score*, *Recall*, dan *Subset Accuracy*, sekaligus menurunkan *Hamming Loss*, menegaskan bahwa pemilihan kombinasi *hyperparameter* yang optimal memainkan peran penting dalam meningkatkan efektivitas klasifikasi multi-label berbasis teks keagamaan.



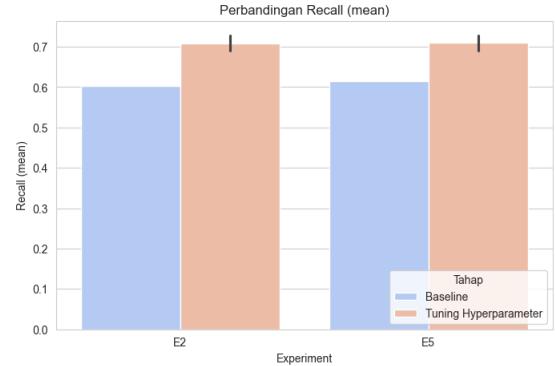
Gambar 1. Perbandingan *F1-Score* Sebelum dan Sesudah *Tuning*



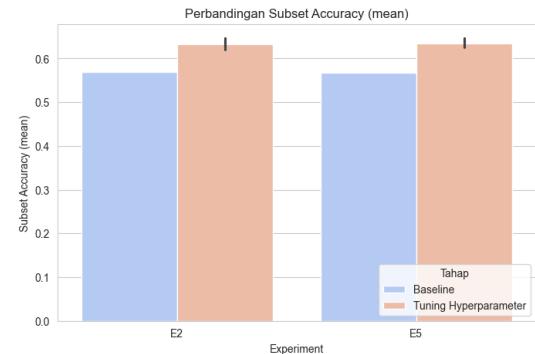
Gambar 2. Perbandingan *Hamming Loss* Sebelum dan Sesudah *Tuning*



Gambar 3. Perbandingan *Precision* Sebelum dan Sesudah *Tuning*



Gambar 4. Perbandingan *Recall* Sebelum dan Sesudah *Tuning*



Gambar 5. Perbandingan *Subset Accuracy* Sebelum dan Sesudah *Tuning*

3.2. Pembahasan

Evaluasi model dilakukan untuk mengukur performa klasifikasi multi-label sebelum dan sesudah *tuning hyperparameter* menggunakan metrik *F1-Score*, *Hamming Loss*, *Precision*, *Recall*, dan *Subset Accuracy*. Hasil evaluasi menunjukkan bahwa *tuning hyperparameter* meningkatkan performa model secara signifikan.

Pada *F1-Score*, model E2 meningkat dari 0.7046 menjadi 0.7789, dan E5 meningkat dari 0.7073 menjadi 0.7734, menunjukkan keseimbangan yang lebih baik antara *Precision* dan *Recall*. *Hamming Loss* menurun, dari 0.0468 menjadi 0.0381 untuk E2, dan 0.0473 menjadi 0.0406 untuk E5, menandakan pengurangan kesalahan klasifikasi. *Precision* pada E2 mengalami sedikit penurunan dari 0.8603 menjadi 0.8459, sedangkan E5 tetap stabil di 0.8239. Namun, *Recall* meningkat secara signifikan, di mana E2 naik dari 0.6723 menjadi 0.7220 dan E5 dari 0.6780 menjadi 0.7461, yang menunjukkan peningkatan kemampuan model dalam mengenali label yang benar.

Selain itu, *Subset Accuracy* juga meningkat, dari 0.6151 menjadi 0.6545 untuk E2, dan dari 0.6212 menjadi 0.6606 untuk E5, mengindikasikan bahwa model lebih sering memberikan prediksi yang benar secara keseluruhan. Dengan peningkatan yang

konsisten di berbagai metrik, *tuning hyperparameter* terbukti meningkatkan kualitas model klasifikasi multi-label.

Evaluasi lebih lanjut dilakukan untuk memahami pengaruh *threshold* dalam klasifikasi multi-label. Tiga nilai *threshold* utama diuji: 0.35, 0.40, dan 0.50. *Threshold* 0.35 memberikan *Recall* tertinggi, menangkap lebih banyak label tetapi meningkatkan false positives. *Threshold* 0.50 lebih selektif dengan *Precision* tertinggi (0.8524), namun meningkatkan false negatives. *Threshold* 0.40 menawarkan keseimbangan optimal dengan *F1-Score* terbaik (0.7758), serta keseimbangan yang lebih baik antara *Precision* (0.8261) dan *Recall* (0.7319), sehingga dipilih sebagai *threshold* optimal dalam penelitian ini.

Konfigurasi *hyperparameter* yang memberikan performa terbaik adalah *batch size* 8, *learning rate* 0.0002, *dropout rate* 0.3, *optimizer* AdamW, *pooling strategy* dynamic, *filter CNN* 256 & 128, dan *kernel size* 5 & 3. Dengan kombinasi ini, model berhasil meningkatkan akurasi klasifikasi multi-label secara signifikan.

Kesimpulan dari hasil evaluasi menunjukkan bahwa *tuning hyperparameter* dan pemilihan *threshold* yang optimal berperan penting dalam meningkatkan performa model, dengan *trade-off* antara cakupan label dan keakuratan prediksi yang perlu disesuaikan dengan kebutuhan klasifikasi. Dengan demikian, pendekatan yang diterapkan dalam penelitian ini berhasil menghasilkan sistem klasifikasi multi-label yang lebih akurat dan efektif dalam menghubungkan teks diskusi keagamaan dengan referensi ayat Al-Qur'an dan Hadits.

4. KESIMPULAN DAN SARAN

Penelitian ini mengusulkan model hibrida BERT-CNN untuk klasifikasi multi-label diskusi keagamaan, dengan tujuan menghubungkan pertanyaan jamaah pada platform digital dengan referensi Al-Qur'an dan Hadits. Hasil penelitian menunjukkan bahwa integrasi BERT sebagai representasi kontekstual dan CNN sebagai ekstraktor pola lokal efektif dalam meningkatkan performa klasifikasi multi-label. Penerapan *tuning hyperparameter* terbukti meningkatkan *F1-Score*, *Recall*, dan *Subset Accuracy* sekaligus menurunkan *Hamming Loss*, yang berarti kesalahan klasifikasi dapat ditekan secara signifikan. Selain itu, pemilihan *threshold* yang tepat (0.40) memberikan keseimbangan terbaik antara *Precision* dan *Recall*, sehingga menghasilkan prediksi yang lebih akurat dan konsisten.

Kontribusi utama penelitian ini meliputi:

1. Mengembangkan arsitektur hibrida BERT-CNN yang terbukti efektif untuk klasifikasi multi-label teks keagamaan berbahasa Indonesia.

2. Menyediakan *dataset* beranotasi manual yang menghubungkan pertanyaan jamaah dengan ayat Al-Qur'an dan Hadits.
3. Menunjukkan pentingnya pemilihan *hyperparameter* dan *threshold* dalam meningkatkan kinerja model multi-label.

Berdasarkan temuan ini, terdapat beberapa arah pengembangan selanjutnya. Pertama, perluasan cakupan *dataset* di luar tema Ramadhan agar model mampu menangani variasi topik yang lebih beragam. Kedua, eksplorasi arsitektur transformer lain, seperti RoBERTa, ALBERT, atau GPT-based models, untuk membandingkan performa dengan pendekatan BERT-CNN. Ketiga, penerapan model pada skenario nyata, misalnya dalam sistem *chatbot* keagamaan atau aplikasi pencarian referensi Al-Qur'an dan Hadits, guna menguji efektivitas model dalam praktik. Keempat, optimasi lebih lanjut melalui teknik *ensemble learning* atau *fine-tuning* dengan *dataset* yang lebih besar untuk meningkatkan generalisasi model.

Dengan rekomendasi tersebut, penelitian selanjutnya diharapkan dapat mengembangkan sistem klasifikasi multi-label teks keagamaan yang lebih akurat, adaptif, dan aplikatif dalam mendukung akses informasi keagamaan berbasis kecerdasan buatan.

DAFTAR PUSTAKA

- AHMADI, H. A., & CHOWANDA, A. 2023. Clickbait Classification Model on Online News With Semantic Similarity Calculation Between News Title and Content. Building of Informatics Technology and Science (Bits), 4(4). <https://doi.org/10.47065/bits.v4i4.3030>
- AHMED, S. BIN. 2022. Sentence Continuation Inference of Urdu Text by BERT Technique. SSRN Electronic Journal, 1(1). <https://doi.org/10.2139/ssrn.4144163>
- AKBAR, I., FAISAL, M., & CHAMIDY, T. 2024. Multi-label classification of Indonesian qur'an translation using long short-term memory model. Computer Network, Computing, Electronics, and Control Journal, 4(3), 119–128. <https://kinetik.umm.ac.id/index.php/kinetik/article/view/1901>
- ALDREABI, E., & BLACKBURN, J. 2023. Enhancing Automated Hate Speech Detection: Addressing Islamophobia and Freedom of Speech in Online Discussions. Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2023, 644–651. <https://doi.org/10.1145/3625007.3627487>
- ALTAMMAMI, S., ATWELL, E., & ALSALKA, A. 2020. Constructing a Bilingual Hadith

- Corpus Using a Segmentation Tool. Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), 3390–3398. <https://aclanthology.org/2020.lrec-1.415>
- ALYOOBI, K. H., ALOTAIBI, F. S., KUMAR, A., GUPTA, V., & SHARMA, A. 2023. A Novel Multi-Layer Feature Fusion-Based BERT-CNN for Sentence Representation Learning and Classification. *Robotic Intelligence and Automation*, 43(6), 704–715. <https://doi.org/10.1108/ria-04-2023-0047>
- ARKOK, B., & ZEKI, A. M. 2021. Classification of Qur'anic topics based on imbalanced classification. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 678–687. <https://doi.org/10.11591/ijeecs.v22.i2.pp678-687>
- ARSLAN, M., & CRUZ, C. 2024. Business Text Classification With Imbalanced Data and Moderately Large Label Spaces for Digital Transformation. *Applied Network Science*, 9(1). <https://doi.org/10.1007/s41109-024-00623-5>
- BUDIMAN, I., FAISAL, M. R., FARIDHAH, A., FARMADI, A., MAZDADI, M. I., SARAGIH, T. H., & ABADI, F. 2024. Classification Performance Comparison of BERT and IndoBERT on SelfReport of COVID-19 Status on Social Media. *Journal of Computer Sciences Institute*, 30, 61–67. <https://doi.org/10.35784/jcsi.5564>
- CAI, L., SHEN, Y., LIU, T., & ZHANG, K. 2020. A Hybrid BERT Model That Incorporates Label Semantics via Adjustive Attention for Multi-Label Text Classification. *Ieee Access*, 8, 152183–152192. <https://doi.org/10.1109/access.2020.3017382>
- CHOIRULFIKRI, M. R., LHAKSAMANA, K. M., & FARABY, S. AL. 2022. A Multi-Label Classification of Al-Quran Verses Using Ensemble Method and Naïve Bayes. *Building of Informatics, Technology and Science (BITS)*, 3(4), 473–479. <https://doi.org/10.47065/bits.v3i4.1287>
- DEVLIN, J., CHANG, M.-W., LEE, K., & TOUTANOVA, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 1(1), 1–16. <https://doi.org/10.48550/arXiv.1810.04805>
- HUTAMA, L. B., & SUHARTONO, D. 2022. Indonesian Hoax News Classification With Multilingual Transformer Model and BERTopic. *Informatica*, 46(8). <https://doi.org/10.31449/inf.v46i8.4336>
- KUSTIAWAN, R., ADIWIJAYA, A., & PURBOLAKSONO, M. D. 2022. A Multi-label Classification on Topic of Hadith Verses in Indonesian Translation using CART and Bagging. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(2), 868. <https://doi.org/10.30865/mib.v6i2.3787>
- LI, Q., XIAO, J. Z., & ZHAO, Y. 2023. Research on the Classification of New Energy Industry Policy Texts Based on BERT Model. *Sustainability*, 15(14), 11186. <https://doi.org/10.3390/su151411186>
- NABIILAH, G. Z., AL FARABY, S., & PURBOLAKSONO, M. D. 2021. Classification of Hadith Topic of Indonesian Translation Using K-Nearest Neighbor and Chi-Square. *Intl. Journal on ICT*, 7(2), 11–22. <https://doi.org/10.34818/ijoict.v7i2.573>
- NABIILAH, G. Z., ALAM, I. N., PURWANTO, E. S., & HIDAYAT, M. F. 2024. Indonesian Multilabel Classification Using IndoBERT Embedding and MBERT Classification. *International Journal of Electrical and Computer Engineering (Ijece)*, 14(1), 1071. <https://doi.org/10.11591/ijece.v14i1.pp1071-1078>
- NAUFAL, M. A., & GIRSANG, A. S. 2024. Traffic Accident Classification Using IndoBERT. *International Journal of Informatics and Communication Technology (Ij-Ict)*, 13(1), 42. <https://doi.org/10.11591/ijict.v13i1.pp42-49>
- NISSA, N. K., & YULANTI, E. 2023. Multi-Label Text Classification of Indonesian Customer Reviews Using Bidirectional Encoder Representations From Transformers Language Model. *International Journal of Electrical and Computer Engineering (Ijece)*, 13(5), 5641. <https://doi.org/10.11591/ijece.v13i5.pp5641-5652>
- NOUH, S. E., & ALSAYAT, A. 2020. The Multi-Class Classification for the First Six Surats of the Holy Quran. *IJACSA) International Journal of Advanced Computer Science and Applications*, 11(1), 327–332. www.ijacs.thesai.org
- NUHA, U., & ROCHMAWATI, N. 2019. Klasifikasi Kesahihan Hadits Berdasarkan Perawi Hadits Menggunakan Principal Component Analysis (PCA) dan Backpropagation Neural Network (BPNN). *Journal of Informatics and Computer Science*, 01(2).
- SYAKHRANI, A. W. 2023. Fungsi, Kedudukan Perbandingan Hadits Dengan Al-Qur'an. *MUSHAF JOURNAL : Jurnal Ilmu Al Quran Dan Hadis*, 3(1), 51–58.
- TORRES, J. N., MORA, M., HERNÁNDEZ-GARCÍA, R., BARRIENTOS, R. J., FREDES, C., & VALENZUELA, A. 2020. A review of convolutional neural network applied to fruit image processing. *Applied*

- Sciences (Switzerland), 10(10).
<https://doi.org/10.3390/app10103443>
- WANG, H., TIAN, K., WU, Z., & WANG, L. 2021. A short text classification method based on convolutional neural network and semantic extension. International Journal of Computational Intelligence Systems, 14(1), 367–375.
<https://doi.org/10.2991/ijcis.d.201207.001>
- WASFEY, A., ELREFAI, E., MUHAMMAD, M., & NAWAZ, H. 2022. Stars at Qur'an QA 2022: Building Automatic Extractive Question Answering Systems for the Holy Qur'an with Transformer Models and Releasing a New Dataset. Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection, 146–153. <https://aclanthology.org/2022.osact-1.18>
- ZHOU, Y., LI, J., CHI, J., TANG, W., & ZHENG, Y. 2022. Set-CNN: A text convolutional neural network based on semantic extension for short text classification. Knowledge-Based Systems, 257(1), 109948.
<https://doi.org/10.1016/J.KNOSYS.2022.109948>