

**CHATBOT PEMBELAJARAN BAHASA ARAB BERBASIS NLP
MENGUNAKAN *FINE-TUNING* ARABERT**

SKRIPSI

**Oleh:
AL HUMAYROH
NIM. 220605110150**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

**CHATBOT PEMBELAJARAN BAHASA ARAB BERBASIS NLP
MENGUNAKAN *FINE-TUNING* ARABERT**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
AL HUMAYROH
NIM. 220605110150

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
202**

HALAMAN PERSETUJUAN

**CHATBOT PEMBELAJARAN BAHASA ARAB BERBASIS NLP
MENGUNAKAN *FINE-TUNING* ARABERT**

SKRIPSI

Oleh:
AL HUMAYROH
NIM. 220605110150

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 20 Mei 2026

Pembimbing I,



Prof. Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

Pembimbing II,



Dr. Ir. Fachrul Kurniawan ST., M.MT., IPU
NIP. 197710202009121001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

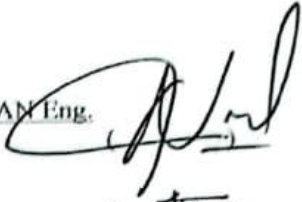
CHATBOT PEMBELAJARAN BAHASA ARAB BERBASIS NLP MENGUNAKAN *FINE-TUNING* ARABERT

SKRIPSI

Oleh :
AL HUMAYROH
NIM. 220605110150

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 20 Mei 2026

Susunan Dewan Penguji

Ketua Penguji	: <u>Dr. Ir. Fresy Nugroho, ST., MT, IPM., ASEAN Eng.</u> NIP. 19710722 201101 1 001	
Anggota Penguji I	: <u>Fatchurrohman, M.Kom</u> NIP. 19700731 200501 1 002	()
Anggota Penguji II	: <u>Prof. Dr. Muhammad Faisal, M.T</u> NIP. 19740510 200501 1 007	()
Anggota Penguji III	: <u>Dr. Ir. Fachrul Kurniawan ST., M.MT., IPU</u> NIP. 197710202009121001	()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Maulana Malik Ibrahim Malang




Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Al Humayroh
NIM : 220605110150
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Chatbot Pembelajaran Bahasa Arab Berbasis NLP
Menggunakan *Fine-Tuning* AraBERT

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 25 Mei 2026
Yang membuat pernyataan,



Al Humayroh
NIM.220605110150

MOTTO

“be the best in anycase but don’t feel the best”

HALAMAN PERSEMBAHAN

“Puji syukur kehadiran Allah *Subhanahu wa Ta’ala* atas segala nikmat, pertolongan, dan kemudahan yang telah diberikan. Sholawat serta salam senantiasa tercurah kepada Nabi Muhammad *Shallallahu ‘Alaihi Wassalam*.”

Dengan penuh rasa syukur *Alhamdulillah*, skripsi ini saya persembahkan kepada kedua orang tua tercinta, Baba Nanang Junaedi dan Mama Pujiastri, yang tidak pernah berhenti mendoakan saya dalam setiap langkah kehidupan. Terima kasih atas segala kasih sayang, dukungan, motivasi, perhatian, serta pengorbanan luar biasa yang telah diberikan dengan penuh ketulusan. Semua perjuangan, kerja keras, dan cinta yang diberikan menjadi kekuatan terbesar bagi saya untuk terus bertahan, belajar, dan menyelesaikan pendidikan hingga tahap ini.

Skripsi ini juga saya persembahkan untuk diri saya sendiri yang telah berusaha bertahan, belajar, melawan rasa lelah dan ragu, serta terus berjuang menyelesaikan seluruh proses perkuliahan hingga berada di tahap ini. Semua proses yang telah dilewati menjadi pelajaran berharga untuk terus berkembang dan tidak mudah menyerah.

Terima kasih kepada dosen pembimbing dan dosen penguji yang telah memberikan ilmu, arahan, kritik, dan saran selama proses penyusunan skripsi ini.

Terima kasih pula kepada sahabat dan teman-teman yang selalu memberikan dukungan, bantuan, doa, dan semangat selama perjalanan perkuliahan.

KATA PENGANTAR

Alhamdulillah, segala puji dan syukur penulis panjatkan kehadirat Allah *Subhanahu wa Ta'ala* atas segala rahmat, nikmat, serta kemudahan yang telah diberikan, sehingga penulis dapat menyelesaikan program studi Teknik Informatika di Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang dengan menyelesaikan skripsi yang berjudul “Chatbot Pembelajaran Bahasa Arab Berbasis NLP Menggunakan *Fine-Tuning* AraBERT”.

Pada penelitian dan penulisan skripsi ini, penulis sadar bahwa dalam penyusunan skripsi ini tak luput dari dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Prof. Dr. Hj Ilfi Nur Diana, M.Si, selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. Muhammad Faisal, M.T selaku dosen pembimbing I dan Dr. Ir. Fachrul Kurniawan ST., M.MT ., IPU selaku dosen pembimbing II yang telah bersedia meluangkan waktunya untuk membimbing dan memberikan arahan dalam penyusunan skripsi ini.
5. Dr. Ir. Fresy Nugroho, ST., MT, IPM., ASEAN Eng. selaku dosen penguji I dan Fatchurrohman, M.Kom selaku dosen penguji II yang telah memberikan saran dan arahan dari seminar proposal hingga sidang skripsi penulis.

6. Seluruh dosen dan jajaran staf Program Studi Teknik Informatika yang telah memberikan ilmu serta yang secara langsung maupun tidak langsung terlibat dalam proses penyusunan skripsi ini.
7. Baba, Mama, serta adik-adik penulis, yang tidak pernah berhenti memberikan semangat, dukungan dan doa yang tiada putusnya serta menghibur penulis sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
8. Teman-teman Teknik Informatika yang telah menemani dan memberikan banyak bantuan kepada saya selama masa perkuliahan.
9. Semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan kontribusi, saran dan dukungan dalam penulisan skripsi ini.

Akhir kata, Penulis menyadari bahwa dalam penyusunan skripsi ini masih jauh dari kesempurnaan, baik dari segi keilmuan dan kepenulisan. Oleh karena itu, penulis berharap semoga skripsi ini diterima sebagai amal ibadah yang bermanfaat. Semoga penulisan skripsi ini dapat bermanfaat bagi banyak pihak.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 25 Mei 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	Error! Bookmark not defined.
HALAMAN PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
ABSTRAK	xv
ABSTRACT	xvi
مستخلص البحث.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
BAB II STUDI PUSTAKA	7
2.1 Penelitian Terkait	7
2.2 Dasar Teori.....	13
2.2.1 Pembelajaran Bahasa Arab	13
2.2.2 Chatbot dalam Pembelajaran Bahasa	14
2.2.3 <i>Natural Language Processing</i> (NLP).....	15
2.2.4 <i>Sequence Labeling</i> dan <i>Token Classification</i> Berbasis Transformer ...	16
2.2.5 AraBERT sebagai Model Bahasa Arab Pra-latih	17
2.2.6 <i>Fine-Tuning</i> Model Bahasa	18
2.2.7 <i>Hyperparameter</i> Pelatihan Model Transformer	19
2.2.8 Perbedaan Domain Qur'anic dan MSA	20
BAB III METODOLOGI PENELITIAN	21
3.1 Desain Penelitian	21
3.2 Data Penelitian	22
3.2.1 Sumber Data	22
3.2.2 Struktur Dataset MASAQ.....	23
3.3 Preprocessing Dataset	23
3.3.1 Penyaringan <i>Morph_type</i>	24
3.3.2 Normalisasi Teks Arab	24
3.3.3 Pembentukan Sequence	25
3.3.4 Tokenisasi dan <i>Label Alignment</i>	25
3.3.5 Representasi Input Model	26
3.3.6 Karakteristik Data setelah <i>Preprocessing</i>	26
3.4 <i>Fine-Tuning</i> Model AraBERT	27
3.4.1 Arsitektur Model AraBERT	28

3.4.2 Fungsi <i>Loss</i>	28
3.4.3 Optimasi parameter.....	28
3.4.4 Konfigurasi <i>Hyperparameter</i> Pelatihan	29
3.4.5 Algoritma <i>Fine-Tuning</i> AraBERT	29
3.5 Evaluasi Model	33
3.5.1 Metrik Evaluasi Model	33
3.5.2 Analisis Generalisasi Domain	34
3.6 Desain Sistem Chatbot Pembelajaran	35
3.6.1 Input Pengguna	37
3.6.2 Identifikasi Bahasa Input	38
3.6.3 Penerjemahan Bahasa Indonesia ke Bahasa Arab	38
3.6.4 Normalisasi Teks Arab	39
3.6.5 Tokenisasi Teks	40
3.6.6 Model AraBERT <i>Fine-Tuned</i>	41
3.6.7 Prediksi Kelas Kata (<i>POS Tagging</i>)	41
3.6.8 Penggabungan <i>Subword</i>	42
3.6.9 Output Sistem Chatbot Pembelajaran.....	42
3.7 Evaluasi Sistem Chatbot Pembelajaran.....	43
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Data Penelitian	45
4.2 Hasil Preprocessing.....	46
4.3 Data Splitting dan Skema Eksperimen	47
4.3.1 Pembagian Data dan <i>Sequence-Level Splitting</i>	47
4.3.2 Skenario Rasio Split Data.....	48
4.3.3 Tujuan dan Desain Eksperimen	50
4.4 Proses <i>Fine-Tuning</i> Model AraBERT	50
4.4.1 Inisialisasi Tokenizer AraBERT.....	51
4.4.2 Tokenisasi dan <i>Label Alignment</i>	52
4.4.3 Representasi Input Model.....	52
4.4.4 Inisialisasi Model <i>Token Classification</i>	53
4.4.5 Konfigurasi Pelatihan dan <i>Hyperparameter</i>	53
4.4.6 Proses Pelatihan dan Pembaruan Bobot	54
4.4.7 <i>Early Stopping</i> dan Pemilihan <i>Checkpoint</i>	55
4.5 Hasil Pelatihan dan Evaluasi Model	57
4.5.1 Hasil Pelatihan Split 70:15:15	57
4.5.2 Hasil Pelatihan Split 80:10:10	58
4.5.3 Hasil Pelatihan Split 90:5:5	59
4.5.4 Analisis Hasil <i>Fine-Tuning</i>	60
4.5.5 Pemilihan Model Terbaik	61
4.6 Evaluasi Model pada Data Uji	62
4.6.1 Metrik Evaluasi	62
4.6.2 Analisis Confusion Matrix	64
4.6.3 Perbandingan dengan Baseline Model	65
4.7 Implementasi Chatbot Pembelajaran Bahasa Arab	68
4.7.1 Tahapan Proses Sistem Chatbot Pembelajaran.....	69
4.7.2 Proses Inferensi dan Post-processing	70

4.7.3 Antarmuka dan Skenario Penggunaan.....	71
4.7.4 Evaluasi Hasil Klasifikasi Chatbot	75
4.7.5 Analisis Domain Qur’anic dan <i>Modern Standard Arabic</i> (MSA).....	86
4.7.6 Keterbatasan Penelitian dan Implementasi Sistem.....	88
4.7.7 Keterkaitan Chatbot dengan Tujuan Penelitian	89
4.8 Integrasi Penelitian dengan Islam	90
4.8.1 Dimensi <i>Ma’allah</i>	91
4.8.2 Dimensi <i>Ma’annas</i>	92
4.8.3 Dimensi <i>Ma’al ‘Alam</i>	94
BAB V PENUTUP	96
5.1 Kesimpulan	96
5.2 Saran	97
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian.....	21
Gambar 3.2 Pipeline <i>Fine-Tuning</i> Model AraBERT	31
Gambar 3.3 Desain Sistem Chatbot	35
Gambar 4.1 Confusion Matrix pada Data Uji	64
Gambar 4.2 Confusion Matrix <i>Baseline Model</i> (Tanpa <i>Fine-Tuning</i>).....	67
Gambar 4.3 Tampilan Antarmuka Chatbot dengan Input Bahasa Indonesia.....	72
Gambar 4.4 Tampilan Antarmuka Chatbot dengan Input Bahasa Arab	73

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu	12
Tabel 3.1 Distribusi Token per Kelas	27
Tabel 3.2 Input dan Output Proses <i>Fine-Tuning</i>	30
Tabel 3.3 Contoh Hasil Prediksi Kelas Kata	43
Tabel 4.1 Contoh <i>Intermediate Dataset</i>	46
Tabel 4.2 Struktur Data Dataset (HuggingFace)	47
Tabel 4.3 Distribusi Data Train, Validation, dan Test pada Setiap Skenario Split	49
Tabel 4.4 Konfigurasi Pelatihan Model	53
Tabel 4.5 Hasil Pelatihan Model Rasio Split 70:15:15	57
Tabel 4.6 Hasil Pelatihan Model Rasio Split 80:10:10	58
Tabel 4.7 Hasil Pelatihan Model Rasio Split 90:5:5	59
Tabel 4.8 Hasil Evaluasi pada Data Uji	62
Tabel 4.9 <i>Classification Report</i> per Kelas	63
Tabel 4.10 Hasil Evaluasi <i>Baseline Model</i> (Tanpa <i>Fine-Tuning</i>)	66
Tabel 4.11 Dasar Kaidah Penentuan Label Aktual Berdasarkan Nahwu	77
Tabel 4.12 Hasil Pengujian Implementasi Chatbot pada Input Bahasa Arab	78
Tabel 4.13 Klasifikasi Jenis Kesalahan	81
Tabel 4.14 Hasil Pengujian Implementasi Chatbot pada Input Bahasa Indonesia	83

ABSTRAK

Al Humayroh. 2026. **Chatbot Pembelajaran Bahasa Arab Berbasis NLP Menggunakan *Fine-Tuning* AraBERT**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Muhammad Faisal, M.T (II) Dr. Ir. Fachrul Kurniawan ST., M.MT., IPU

Kata kunci: Chatbot, Bahasa Arab, *Natural Language Processing*, *Fine-Tuning*, AraBERT, *Token Classification*.

Penelitian ini bertujuan untuk mengembangkan chatbot pembelajaran bahasa Arab berbasis *Natural Language Processing* menggunakan metode *fine-tuning* AraBERT untuk klasifikasi jenis kata bahasa Arab berupa isim, fi'il, dan harf. Sistem dikembangkan menggunakan dataset *Morphologically and Syntactically Annotated Quran Dataset (MASAQ)* dan diimplementasikan ke dalam chatbot interaktif berbasis web. Hasil implementasi menunjukkan bahwa chatbot mampu melakukan analisis jenis kata bahasa Arab secara otomatis pada kalimat bahasa Arab baku (*Modern Standard Arabic/MSA*) secara cukup konsisten. Hasil evaluasi model menunjukkan nilai *F1-score* sebesar 0.990538 pada proses klasifikasi token bahasa Arab. Selain itu, hasil evaluasi implementasi chatbot pada 15 kalimat bahasa Arab menunjukkan tingkat akurasi klasifikasi token sebesar 92.54% dengan 62 token berhasil diklasifikasikan secara benar dari total 67 token yang diuji. Kesalahan klasifikasi umumnya terjadi pada token berbentuk klitik dan partikel pendek seperti huruf sambung serta prefiks yang melekat pada kata dasar. Secara keseluruhan, sistem mampu membantu proses pembelajaran dasar bahasa Arab serta memberikan analisis klasifikasi kata secara otomatis dan interaktif.

ABSTRACT

Al Humayroh. 2026. *Arabic Learning Chatbot Based on Natural Language Processing Using AraBERT Fine-Tuning*. Undergraduate Thesis. Informatics Engineering Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: (I) Prof. Dr. Muhammad Faisal, M.T (II) Dr. Ir. Fachrul Kurniawan ST., M.MT ., IPU

Keywords: Chatbot, Arabic Language, *Natural Language Processing*, *Fine-Tuning*, AraBERT, *Token Classification*.

This research aims to develop an Arabic learning chatbot based on *Natural Language Processing* using the AraBERT fine-tuning method for classifying Arabic word categories into *isim*, *fi'il*, and *harf*. The system was developed using the *Morphologically and Syntactically Annotated Quran Dataset (MASAQ)* and implemented into an interactive web-based chatbot. The implementation results show that the chatbot is capable of automatically analyzing Arabic word categories in *Modern Standard Arabic (MSA)* sentences consistently. The model evaluation results achieved an *F1-score* of 0.990538 in Arabic token classification tasks. In addition, the chatbot implementation evaluation on 15 Arabic sentences obtained a token classification accuracy of 92.54%, with 62 correctly classified tokens out of 67 tested tokens. Classification errors generally occurred in clitic tokens and short particles such as conjunctions and prefixes attached to root words. Overall, the system is able to support basic Arabic language learning and provide automatic and interactive Arabic word classification analysis.

مستخلص البحث

الحُمَيَّرَاء. ٢٠٢٦. روبوت محادثة لتعليم اللغة العربية قائم على معالجة اللغة الطبيعية باستخدام الضبط الدقيق لنموذج AraBERT. بحث تخرج. قسم تقنية المعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: الأستاذ الدكتور محمد فيصل، الماجستير. المشرف الثاني: الدكتور المهندس فخر الكورنياوان، الماجستير، IPU.

الكلمات المفتاحية: روبوت المحادثة، اللغة العربية، معالجة اللغة الطبيعية، الضبط الدقيق، AraBERT، تصنيف الكلمات.

يهدف هذا البحث إلى تطوير روبوت محادثة لتعلم اللغة العربية قائم على معالجة اللغة الطبيعية باستخدام أسلوب الضبط الدقيق لنموذج AraBERT لتصنيف أنواع الكلمات العربية إلى اسم وفعل وحرف. تم تطوير النظام باستخدام مجموعة بيانات MASAQ الخاصة بالتحليل الصرفي والنحوي للغة العربية، ثم تم دمجها في روبوت محادثة تفاعلي قائم على الويب. أظهرت نتائج التطبيق أن روبوت المحادثة قادر على تحليل أنواع الكلمات العربية تلقائيًا في جمل اللغة العربية الفصحى الحديثة (MSA) بشكل متسق إلى حدٍ كبير. كما حقق النموذج قيمة $F1-score$ بلغت 0.990538 في مهمة تصنيف الكلمات العربية. بالإضافة إلى ذلك، أظهرت نتائج تقييم تطبيق روبوت المحادثة على 15 جملة عربية نسبة دقة بلغت 92.54٪، حيث تم تصنيف 62 كلمة بشكل صحيح من أصل 67 كلمة تم اختبارها. وظهرت أخطاء التصنيف غالبًا في الكلمات القصيرة واللواحق النحوية مثل حروف العطف والبادئات المتصلة بالكلمات الأساسية. وبشكل عام، يستطيع النظام دعم تعلم اللغة العربية الأساسية وتقديم تحليل تلقائي وتفاعلي لتصنيف الكلمات العربية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahasa Arab merupakan bahasa sumber ajaran Islam yang memiliki peran penting dalam memahami Al-Qur'an dan Hadis. Pemahaman terhadap teks tersebut tidak dapat dilepaskan dari penguasaan kaidah bahasa Arab, terutama dalam mengenali struktur dasar seperti jenis kata (*isim*, *fi'il*, dan *harf*). Allah Ta'ala berfirman:

إِنَّا أَنْزَلْنَاهُ قُرْآنًا عَرَبِيًّا لَعَلَّكُمْ تَعْقِلُونَ

“Sesungguhnya Kami menurunkannya sebagai Qur'an berbahasa Arab agar kamu memahaminya.” (QS. Yusuf: 2).

Ayat tersebut dijelaskan dalam tafsir Ibnu Katsir bahwa Al-Qur'an diturunkan dalam bahasa Arab yang paling fasih dan jelas, sehingga pemahaman kaidah bahasanya memiliki peranan penting dalam menggali makna secara tepat (Ibnu Katsir, 2003a). Hal ini menunjukkan bahwa kemampuan mengenali jenis kata seperti isim, fi'il, dan harf menjadi tahap dasar dalam pembelajaran bahasa Arab. Identifikasi kelas kata ini menjadi fondasi untuk memahami struktur kalimat sebelum memasuki kajian sintaksis yang lebih kompleks (Zaid et al., 2024).

Dalam praktik pembelajaran, pengenalan jenis kata masih sering dilakukan secara manual, sehingga pembelajar kesulitan memperoleh umpan balik cepat dan konsisten. Perkembangan teknologi seperti *Natural Language Processing* (NLP), membuka peluang penerapan model kecerdasan buatan untuk membantu analisis

bahasa Arab secara otomatis menggunakan model kecerdasan buatan (Moradi et al., 2022). Salah satu model yang banyak digunakan adalah AraBERT, yaitu model transformer yang dirancang khusus untuk Bahasa Arab dan mampu melakukan tugas seperti klasifikasi token serta analisis morfologi (El-Alami et al., 2022).

Ketersediaan dataset *Qur'anic Arabic* (MASAQ) menjadikan proses pelatihan model lebih terarah karena dataset tersebut menyediakan anotasi kelas kata pada tingkat token. Pemanfaatan dataset dan model tersebut dalam konteks pembelajaran interaktif masih belum banyak dilakukan. Beberapa penelitian sebelumnya berfokus pada peningkatan relevansi respons chatbot (Alruqi & Alzahrani, 2023), pengembangan model AraBERT untuk pemrosesan bahasa Arab (Antoun et al., 2020), atau pembuatan chatbot edukasi tanpa *transformer* (Alazzam et al., 2023). Terdapat pula penelitian terkait chatbot untuk pembelajaran seperti Mondly (Faroid et al., 2025), serta evaluasi *fine-tuning* model BERT Arab untuk klasifikasi teks (Al-harbi & Algheith, 2025), dan ekstraksi embedding menggunakan AraBERT (Abo-Elghit et al., 2022).

Penelitian-penelitian tersebut umumnya belum diarahkan pada analisis linguistik tingkat token, khususnya dalam mengidentifikasi kelas kata sebagai aspek morfologi dasar dalam konteks pembelajaran bahasa Arab. Sebagian besar studi masih berfokus pada pemrosesan tingkat kalimat atau dokumen serta respons percakapan, sehingga belum secara langsung mendukung pemahaman struktur bahasa pada tingkat dasar (Alazzam et al., 2023; Al-harbi & Algheith, 2025; Alruqi & Alzahrani, 2023). Kondisi ini menunjukkan adanya celah penelitian dalam pemanfaatan *token-level classification* untuk mendukung pembelajaran bahasa

Arab berbasis struktur *linguistik*. Penelitian ini mengisi celah tersebut dengan mengusulkan pengembangan model *fine-tuning* AraBERT untuk klasifikasi kelas kata dasar (*isim*, *fi'il*, dan *harf*) yang diintegrasikan dalam sistem chatbot pembelajaran. Dataset *Qur'anic Arabic* (MASAQ) digunakan sebagai data pelatihan karena memiliki kualitas anotasi linguistik yang konsisten, dengan pertimbangan pada kualitas *corpus*, bukan konteks keagamaan teks.

Pada tahap implementasi, terdapat perbedaan karakteristik antara data pelatihan berbasis dataset *Qur'anic Arabic* dan penggunaan pada teks bahasa Arab baku modern (*Modern Standard Arabic/MSA*). Perbedaan ini berpotensi menimbulkan domain shift yang memengaruhi performa model. Fokus penelitian diarahkan pada kemampuan generalisasi model dalam melakukan klasifikasi kelas kata pada variasi teks yang berbeda, bukan pada klaim performa optimal di seluruh domain.

Sistem yang dikembangkan juga dirancang untuk menerima input dalam bahasa Arab maupun bahasa Indonesia sebagai bentuk adaptasi terhadap pengguna pemula. Evaluasi utama tetap dilakukan pada teks bahasa Arab untuk menjaga validitas pengukuran performa model. Pendekatan ini diharapkan dapat memberikan kontribusi dalam pengembangan media pembelajaran bahasa Arab berbasis NLP yang berfokus pada analisis struktur linguistik dasar secara otomatis.

1.2 Rumusan Masalah

Bagaimana mengembangkan dan mengimplementasikan model *fine-tuning* AraBERT untuk klasifikasi kelas kata bahasa Arab dalam sistem chatbot pembelajaran?

1.3 Batasan Masalah

Agar penelitian lebih terarah, batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan terbatas pada dataset MASAQ berbasis anotasi morfologi ayat-ayat Al-Qur'an karena menyediakan struktur anotasi token-level yang lengkap, konsisten, dan sesuai untuk tugas token classification, khususnya pada label *morph_tag* (jenis kata: *isim*, *fi'il*, dan *harf*).
2. Penelitian ini tidak mencakup:
 - a. Analisis fungsi *sintaksis* (*syntactic_role*)
 - b. *I'rab* atau *case/mood*
 - c. Penerjemahan ayat
 - d. Tafsir Al-Qur'an
 - e. Analisis tajwid
3. Fine-tuning model dilakukan untuk tugas *token classification* (*sequence labeling*), bukan untuk model generatif atau pemahaman makna ayat secara kontekstual.
4. Chatbot hanya memberikan analisis *linguistik* pada tingkat jenis kata, dan tidak menjawab pertanyaan keagamaan yang bersifat fikih, akidah, maupun hukum syari'at.
5. Input pengguna dapat berupa teks bahasa Arab maupun bahasa Indonesia. Input bahasa Indonesia akan diproses melalui tahap penerjemahan ke dalam bahasa Arab sebelum dilakukan klasifikasi.

6. Penelitian ini tidak melakukan evaluasi lintas domain skala besar menggunakan korpus *Modern Standard Arabic* (MSA) sebagai dataset pelatihan utama karena keterbatasan ketersediaan dataset MSA terbuka dengan anotasi token-level morfologi yang setara. Pengujian pada kalimat MSA dilakukan secara terbatas sebagai evaluasi implementatif untuk mengamati kemampuan generalisasi model di luar domain pelatihan.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini, sebagai berikut:

1. Mengembangkan model NLP melalui *fine-tuning* AraBERT untuk melakukan klasifikasi jenis kata (*isim*, *fi'il*, dan *harf*) sebagai fitur *morfologi* dasar bahasa Arab berdasarkan dataset anotasi *morfologi* ayat-ayat Al-Qur'an.
2. Menghasilkan chatbot pembelajaran bahasa Arab yang mampu:
 - a. Menerima input teks bahasa Arab maupun bahasa Indonesia
 - b. Melakukan penerjemahan dari bahasa Indonesia ke bahasa Arab (jika diperlukan)
 - c. Menampilkan hasil klasifikasi jenis kata
 - d. Memberikan penjelasan singkat yang mudah dipahami oleh pengguna
3. Menyediakan media pembelajaran yang membantu pemula mengenali dan memahami jenis kata dalam bahasa Arab secara mandiri.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini yaitu sebagai berikut:

1. Secara Akademis: Memberikan kontribusi pada pengembangan teknologi NLP untuk bahasa Arab, khususnya dalam penerapan *fine-tuning* AraBERT untuk tugas klasifikasi jenis kata berbasis anotasi morfologi Al-Qur'an. Hasil penelitian ini dapat menjadi rujukan bagi pengembangan sistem pembelajaran bahasa Arab berbasis AI.
2. Secara Praktis: Chatbot yang dikembangkan diharapkan dapat membantu pengguna dalam mempelajari bahasa Arab dasar secara interaktif, cepat, dan mudah diakses, khususnya dalam mengenali jenis kata Arab, sehingga mempercepat proses pemahaman morfologi dasar.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Perkembangan chatbot dalam pendidikan semakin pesat, terutama dalam pembelajaran bahasa yang memanfaatkan NLP untuk menganalisis struktur kalimat dan kosakata secara otomatis. Penggunaan model transformer seperti AraBERT memungkinkan pemrosesan linguistik yang lebih akurat melalui fine-tuning pada dataset beranotasi seperti MASAQ. Dalam pembelajaran bahasa Arab, pendekatan ini tidak hanya mendukung identifikasi kelas kata dan struktur gramatikal, tetapi juga dapat diimplementasikan pada chatbot yang memberi umpan balik langsung kepada pengguna. Dengan demikian, NLP berbasis transformer berperan penting dalam pengembangan chatbot yang mampu memberikan analisis gramatikal, bukan sekadar respons percakapan.

Penelitian yang dilakukan oleh (Alruqi & Alzahrani, 2023) mengembangkan chatbot bahasa Arab berbasis *extractive question answering* dengan pendekatan transfer learning menggunakan model transformer bahasa Arab. Dalam penelitian tersebut, AraBERT digunakan sebagai model pralatih yang kemudian di-*fine-tuning* untuk memahami konteks pertanyaan dan jawaban dalam bahasa Arab secara lebih akurat. Hasil penelitian menunjukkan bahwa *fine-tuning* model transformer bahasa Arab, khususnya AraBERT, mampu meningkatkan relevansi dan kualitas jawaban pada sistem chatbot bahasa Arab. Temuan ini menunjukkan bahwa AraBERT efektif digunakan dalam pengembangan sistem NLP bahasa Arab berbasis pemahaman konteks. Namun, penelitian tersebut berfokus pada question answering

dan pencarian jawaban otomatis, sehingga belum mengarah pada fungsi edukatif seperti analisis linguistik tingkat token, identifikasi kelas kata (*isim, fi'il, harf*), maupun evaluasi struktur gramatikal bahasa Arab untuk pembelajaran. Oleh karena itu, penelitian ini mengembangkan pemanfaatan AraBERT ke arah chatbot pembelajaran bahasa Arab yang tidak hanya menjawab, tetapi juga menganalisis struktur kebahasaan secara otomatis.

Penelitian oleh (Antoun et al., 2020) berfokus pada pengembangan AraBERT, yaitu model transformer berbasis BERT yang dipralatih secara khusus untuk bahasa Arab guna meningkatkan performa berbagai tugas *Natural Language Understanding*. Model ini dilatih menggunakan korpus bahasa Arab berskala besar dan dievaluasi pada beberapa tugas NLP, antara lain *sentiment analysis*, *named entity recognition*, dan *question answering*. Hasil penelitian menunjukkan bahwa AraBERT berhasil mencapai atau melampaui *state-of-the-art performance* pada sebagian besar dataset pengujian dan secara konsisten mengungguli model *multilingual BERT*, terutama pada tugas NER dengan peningkatan nilai F1 yang signifikan. Meskipun demikian, penelitian ini berfokus pada pengembangan dan evaluasi model bahasa secara umum serta belum mengkaji penerapan AraBERT pada sistem chatbot pembelajaran bahasa Arab maupun pemanfaatannya untuk analisis gramatikal seperti identifikasi kelas kata dan pemberian umpan balik linguistik secara langsung kepada pengguna.

(Alazzam et al., 2023) mengembangkan chatbot edukasi berbahasa Arab yang dirancang untuk membantu pengguna memperoleh informasi akademik dan regulasi pendidikan melalui interaksi tanya jawab. Sistem chatbot tersebut

dibangun menggunakan pendekatan neural network konvensional untuk memetakan pertanyaan pengguna ke jawaban yang relevan berdasarkan basis pengetahuan yang telah disiapkan. Hasil penelitian menunjukkan bahwa chatbot mampu memberikan respons yang sesuai dengan kebutuhan pengguna dan berfungsi secara efektif sebagai media bantu informasi pendidikan. Meskipun demikian, pendekatan yang digunakan belum memanfaatkan model transformer maupun proses *fine-tuning* pada model bahasa Arab pra-latih, sehingga kemampuan sistem dalam memahami struktur sintaksis bahasa Arab secara mendalam, seperti analisis kelas kata atau struktur gramatikal, masih terbatas dan belum mendukung pemberian umpan balik linguistik secara otomatis dalam konteks pembelajaran bahasa Arab.

Penelitian yang dilakukan oleh (Faroid et al., 2025) mengkaji pemanfaatan chatbot pada aplikasi pembelajaran bahasa Arab sebagai media pembelajaran digital yang interaktif. Hasil kajian menunjukkan bahwa penggunaan chatbot mampu meningkatkan motivasi belajar, memperkaya kosakata, serta mendukung latihan percakapan bahasa Arab secara lebih aktif dan kontekstual. Temuan ini memperlihatkan bahwa chatbot memiliki potensi signifikan sebagai sarana pembelajaran bahasa Arab di era digital. Namun, penelitian tersebut lebih berfokus pada aspek pedagogis dan pengalaman pengguna dalam penggunaan chatbot sebagai media belajar, tanpa membahas mekanisme teknis pemrosesan bahasa alami yang mendasari sistem, serta belum mengkaji pengembangan model NLP seperti *fine-tuning* AraBERT untuk analisis linguistik otomatis, termasuk identifikasi kelas kata dan struktur gramatikal bahasa Arab. Oleh karena itu,

penelitian ini melengkapi aspek tersebut melalui pengembangan chatbot pembelajaran bahasa Arab berbasis NLP yang tidak hanya interaktif, tetapi juga mampu memberikan analisis gramatikal secara otomatis.

Selanjutnya, penelitian (Al-harbi & Algheith, 2025) mengeksplorasi fine-tuning berbagai model BERT bahasa Arab dan multilingual untuk tugas klasifikasi teks kriminal berbahasa Arab. Penelitian ini membangun dataset cuitan dari platform X (Twitter), melakukan pra-pemrosesan, penyeimbangan data, serta evaluasi terhadap beberapa model pralatih. Hasil penelitian menunjukkan bahwa model BERT khusus bahasa Arab memberikan performa lebih baik dibandingkan model multilingual dalam memahami konteks bahasa Arab pada tugas klasifikasi teks. Temuan ini memperkuat bahwa fine-tuning model pralatih bahasa Arab efektif untuk meningkatkan performa NLP berbahasa Arab. Namun, penelitian tersebut berfokus pada klasifikasi tingkat dokumen dalam domain kriminal dan belum mengkaji analisis linguistik tingkat token, seperti identifikasi kelas kata (isim, fi'il, harf) atau struktur gramatikal kalimat Arab, serta belum diimplementasikan dalam chatbot pembelajaran bahasa Arab berbasis umpan balik gramatikal otomatis. Oleh karena itu, penelitian ini mengisi celah tersebut melalui fine-tuning AraBERT untuk analisis morfosintaksis bahasa Arab dalam sistem chatbot edukatif.

Penelitian oleh (Abo-Elghit et al., 2022) berfokus pada pemanfaatan model pralatih AraBERT untuk ekstraksi embedding teks Arab dalam tugas *semantic textual similarity*, khususnya dengan membandingkan penggunaan AraBERT sebagai *feature extractor* dan sebagai model yang di-*fine-tuning*. Penelitian ini mengombinasikan embedding kontekstual AraBERT dengan berbagai skema

representasi teks lain serta model regresi untuk memprediksi skor relevansi antar pasangan teks Arab, dan hasil eksperimen menunjukkan bahwa perbedaan kinerja antara AraBERT yang digunakan sebagai *feature extractor* dan yang di-*fine-tuning* relatif kecil, dengan beberapa varian AraBERT-Large bahkan unggul tanpa proses *fine-tuning* penuh. Meskipun demikian, penelitian tersebut masih berfokus pada tingkat kesamaan semantik antarkalimat secara umum dan belum mengeksplorasi penerapan *fine-tuning* AraBERT pada tugas token-level classification yang lebih linguistik, seperti klasifikasi morfologi atau nahwu kata per kata, serta belum mengintegrasikan hasil model ke dalam sistem pembelajaran interaktif. Oleh karena itu, celah penelitian ini dapat diisi dengan menerapkan *fine-tuning* AraBERT secara khusus untuk tugas klasifikasi morfologi bahasa Arab berbasis token dan mengimplementasikannya dalam bentuk chatbot edukatif yang memberikan umpan balik gramatikal secara langsung kepada pengguna

Berdasarkan penelitian tersebut, pemanfaatan chatbot dalam pendidikan telah dilakukan dengan berbagai pendekatan, mulai dari pemanfaatan model transformer untuk question answering hingga penggunaan chatbot sebagai media pendukung pembelajaran kosakata dan percakapan. Selain itu, model transformer berbahasa Arab seperti AraBERT juga telah banyak diteliti dan terbukti memberikan performa yang unggul pada berbagai tugas NLP. Untuk memberikan gambaran yang lebih sistematis mengenai fokus penelitian, hasil utama, serta keterbatasan masing-masing studi, maka perbandingan penelitian terdahulu disajikan dalam Tabel 2.1.

Tabel 2.1 Perbandingan Penelitian Terdahulu

No	Peneliti	Judul	Hasil Utama	Celah Penelitian
1	Alruqi & Alzahrani (2023)	Evaluation of an Arabic Chatbot Based on Extractive Question-Answering Transfer Learning and Language Transformers	Transformer meningkatkan relevansi jawaban dan pemahaman konteks	Tidak difokuskan pada pembelajaran bahasa Arab dan analisis kelas kata dasar bahasa Arab
2	Antoun et al. (2020)	AraBERT: Transformer-based Model for Arabic Language Understanding	Model pralatih Arab menghasilkan performa tinggi pada tugas NLP	Belum diimplementasikan dalam chatbot pembelajaran atau analisis kelas kata
3	Alazzam et al. (2023)	Arabic Educational Neural Network Chatbot	Membantu pengguna memperoleh informasi regulasi pendidikan	Tidak menggunakan transformer / fine-tuning dan tidak menganalisis struktur kalimat
4	Faroid et al. (2023)	Chatbot Mondly Sebagai Media Inovatif Dalam Pembelajaran Bahasa Arab	Mendukung latihan kosakata dan percakapan	Fokus pada penggunaan aplikasi, belum mengembangkan NLP atau analisis nahwu
5	Al-Harbi & Al-Ghieth (2025)	Fine-Tuning Arabic and Multilingual BERT Models for Crime Classification to Support Law Enforcement and Crime Prevention	Model Arab outperform multilingual	Tidak diarahkan pada chatbot pembelajaran atau token classification gramatikal
6	Abo-Elghit et al. (2022)	Embedding Extraction for Arabic Text Using the AraBERT Model	Model memahami relasi semantik token	Tidak diterapkan untuk pembelajaran bahasa atau analisis kelas kata/token

Berdasarkan perbandingan pada Tabel 2.1, terlihat bahwa sebagian besar penelitian terdahulu masih berfokus pada pemrosesan bahasa Arab pada tingkat dokumen atau kalimat, serta pemanfaatan chatbot sebagai media interaksi tanpa analisis linguistik mendalam. Sehingga penelitian ini mengisi celah penelitian dengan menerapkan fine-tuning AraBERT untuk token classification dan mengintegrasikannya ke dalam chatbot pembelajaran bahasa Arab.

2.2 Dasar Teori

Dasar teori merupakan landasan konseptual yang digunakan untuk mendukung penelitian serta menjelaskan keterkaitan antar konsep yang digunakan dalam pengembangan sistem. Pada penelitian ini, dasar teori mencakup konsep pembelajaran bahasa Arab, ilmu nahwu, Natural Language Processing (NLP), model Transformer, AraBERT, token classification, serta chatbot pembelajaran yang menjadi dasar dalam pembangunan sistem klasifikasi kelas kata bahasa Arab (Mulyadi, 2013).

2.2.1 Pembelajaran Bahasa Arab

Pembelajaran bahasa Arab bertujuan membekali peserta didik dengan kemampuan memahami struktur dan kaidah bahasa Arab sebagai dasar dalam memahami teks berbahasa Arab, khususnya ayat-ayat Al-Qur'an (Hanif, 2025). Pemahaman tersebut tidak hanya berkaitan dengan penguasaan kosakata, tetapi juga kemampuan menganalisis fungsi kata dalam susunan kalimat.

Ilmu nahwu merupakan cabang utama yang berperan dalam analisis struktur kalimat bahasa Arab (Shohib & Fuadah, 2022). Ilmu ini membahas kedudukan dan fungsi kata, seperti isim, fi'il, dan harf, serta hubungan sintaktis antar kata. Pemahaman nahwu penting karena perbedaan fungsi kata dapat memengaruhi struktur dan makna gramatikal suatu kalimat (Roziqi & Bakar, 2025).

Dalam pembelajaran nahwu, pengenalan jenis kata dan konsep i'rab menjadi dasar utama analisis gramatikal. I'rab berfungsi sebagai penanda hubungan sintaktis antar kata dalam kalimat (Rini, 2019). Oleh karena itu, diperlukan media pembelajaran yang mampu menyajikan analisis gramatikal secara sistematis.

Dalam konteks penelitian ini, analisis gramatikal nahwu difokuskan pada pengenalan kelas kata dasar dan direalisasikan secara teknis melalui tugas token classification, bukan sebagai pengganti kajian nahwu secara menyeluruh.

2.2.2 Chatbot dalam Pembelajaran Bahasa

Chatbot merupakan sistem berbasis komputer yang dirancang untuk berinteraksi dengan pengguna menggunakan bahasa alami. Dalam konteks pendidikan bahasa, chatbot berfungsi sebagai media pembelajaran interaktif yang mampu menyampaikan materi, menjawab pertanyaan, serta memberikan umpan balik secara otomatis kepada pengguna (AlHumoud et al., 2018). Kehadiran chatbot memungkinkan proses pembelajaran berlangsung secara mandiri dan berkelanjutan melalui interaksi berbasis teks.

Chatbot dalam pembelajaran bahasa dapat dibedakan menjadi chatbot berbasis aturan (*rule-based*) dan chatbot berbasis *Natural Language Processing* (NLP). Chatbot berbasis aturan bekerja dengan mencocokkan input pengguna terhadap pola atau aturan yang telah ditentukan sebelumnya. Pendekatan ini memiliki keterbatasan dalam menangani variasi bentuk kata, perbedaan susunan kalimat, serta ambiguitas fungsi kata, sehingga respons yang dihasilkan cenderung kaku dan bergantung pada kelengkapan aturan yang dirancang (Lalwani et al., 2018). Dalam pembelajaran bahasa Arab, keterbatasan ini menjadi semakin signifikan karena satu kata dapat memiliki fungsi gramatikal yang berbeda bergantung pada konteks kalimat.

Sedangkan chatbot berbasis NLP memanfaatkan model bahasa untuk memahami makna dan konteks input pengguna secara lebih fleksibel. Pendekatan

ini memungkinkan chatbot menganalisis struktur bahasa dan memberikan respons berdasarkan informasi linguistik yang terkandung dalam setiap kata (Alazzam et al., 2023). Hal tersebut menjadikan chatbot berbasis NLP lebih relevan untuk pembelajaran bahasa Arab yang memiliki struktur gramatikal kompleks, khususnya dalam membantu pengguna memahami fungsi kata seperti *isim*, *fi'il*, dan *harf* melalui interaksi langsung. Oleh karena itu, penerapan NLP dalam chatbot pembelajaran bahasa Arab dipandang lebih efektif dibandingkan pendekatan berbasis aturan.

2.2.3 *Natural Language Processing (NLP)*

Natural Language Processing (NLP) merupakan pendekatan komputasional yang digunakan untuk menganalisis dan memahami teks berbahasa Arab, khususnya dalam mengidentifikasi struktur gramatikal dan fungsi kata dalam kalimat. Pendekatan ini memungkinkan sistem memproses bahasa Arab secara kontekstual, bukan sekadar berdasarkan aturan statis (Badawi, 2021).

Bahasa Arab memiliki kompleksitas linguistik yang tinggi, ditandai dengan morfologi yang kaya, penggunaan imbuhan dan klitik, serta perubahan bentuk kata yang bergantung pada konteks gramatikal. Kompleksitas tersebut menyulitkan pendekatan berbasis aturan untuk melakukan analisis bahasa secara akurat, sehingga pendekatan NLP berbasis pembelajaran data menjadi lebih relevan dalam menangkap variasi bentuk dan fungsi kata dalam teks bahasa Arab.

Dalam penelitian ini, NLP diterapkan melalui tugas *sequence labeling*, yaitu proses pemberian label gramatikal pada setiap token dalam suatu urutan teks berdasarkan konteksnya. Pendekatan ini sejalan dengan kebutuhan analisis jenis

kata dan fungsi sintaktis dalam kalimat bahasa Arab, sehingga mendukung penerapan model bahasa berbasis *transformer* untuk analisis struktur gramatikal.

2.2.4 *Sequence Labeling dan Token Classification Berbasis Transformer*

Sequence labeling merupakan salah satu tugas dalam *Natural Language Processing* (NLP) yang bertujuan memberikan label pada setiap token dalam suatu urutan teks dengan mempertimbangkan konteks keseluruhan kalimat (Nguyen & Guo, 2007). Tugas ini umum digunakan dalam *Part-of-Speech (POS) tagging*, *Named Entity Recognition (NER)*, serta analisis struktur gramatikal, karena memungkinkan identifikasi fungsi kata secara kontekstual dalam sebuah kalimat. Dalam konteks penelitian ini, *sequence labeling* digunakan untuk mengklasifikasikan kelas kata bahasa Arab ke dalam kategori Isim, Fi'il, dan Harf berdasarkan konteks kemunculannya dalam kalimat.

Dalam implementasi modern NLP, pendekatan *sequence labeling* direalisasikan melalui skema *token classification*, yaitu proses pemberian label pada setiap token secara individual dengan memanfaatkan representasi konteks yang dihasilkan oleh model bahasa. Berbeda dengan *sequence classification* yang memberikan satu label untuk keseluruhan teks, *token classification* bekerja pada tingkat token sehingga lebih sesuai untuk tugas analisis sintaktis yang membutuhkan identifikasi peran kata secara spesifik (Jafari, 2022).

Perkembangan metode *sequence labeling* mengalami evolusi dari pendekatan berbasis fitur manual menuju model pembelajaran representasi otomatis. Model Conditional Random Field (CRF) banyak digunakan karena mampu menjaga konsistensi antar label dalam suatu urutan, namun sangat

bergantung pada perancangan fitur secara manual. Selanjutnya, model berbasis Bidirectional Long Short-Term Memory (BiLSTM) meningkatkan kemampuan representasi konteks dua arah, tetapi masih memiliki keterbatasan dalam efisiensi komputasi dan pemodelan relasi jarak jauh (Ma & Hovy, 2016).

Transformer merupakan arsitektur deep learning yang diperkenalkan oleh (Vaswani et al., 2017) dengan mekanisme *self-attention* untuk memodelkan hubungan antar token dalam suatu sequence secara paralel. Berbeda dengan pendekatan recurrent seperti LSTM, Transformer mampu menangkap konteks global secara lebih efisien dan efektif pada *sequence* panjang.

Arsitektur Transformer kemudian diperkenalkan dengan mekanisme *self-attention* yang memungkinkan pemodelan relasi antar token secara paralel dan menyeluruh dalam satu waktu (Gong et al., 2020). Kemampuan ini menjadikan Transformer lebih efektif dalam menangkap konteks global dibandingkan pendekatan sebelumnya (Antoun et al., 2020). Dalam penelitian ini, model berbasis Transformer dipilih karena kemampuannya menghasilkan representasi kontekstual yang lebih kaya, sehingga lebih sesuai untuk tugas klasifikasi kelas kata bahasa Arab yang memiliki kompleksitas morfologis dan sintaktis tinggi.

2.2.5 AraBERT sebagai Model Bahasa Arab Pra-latih

Berdasarkan pemaparan mengenai model berbasis Transformer pada subbab sebelumnya, penelitian ini mengimplementasikan model pra-latih khusus bahasa Arab, yaitu AraBERT. AraBERT merupakan model bahasa pra-latih (*pretrained language model*) yang dikembangkan secara khusus untuk memproses teks berbahasa Arab. Model ini dibangun berdasarkan arsitektur Transformer dan dilatih

menggunakan korpus bahasa Arab berskala besar, sehingga mampu mempelajari representasi linguistik bahasa Arab secara kontekstual (Antoun et al., 2020).

Dibandingkan model bahasa umum, AraBERT dirancang dengan mempertimbangkan karakteristik khas bahasa Arab, seperti morfologi yang kaya, variasi bentuk kata, serta penggunaan imbuhan dan klitik (El Moubtahij et al., 2022). Proses pra-pelatihan pada data bahasa Arab memungkinkan AraBERT menghasilkan representasi token yang lebih sesuai untuk menangkap makna dan fungsi kata dalam konteks kalimat. Hal ini menjadikan AraBERT lebih unggul dibandingkan model multibahasa atau model yang tidak dilatih secara khusus pada bahasa Arab.

Dalam konteks penelitian ini, AraBERT digunakan untuk mendukung tugas *token classification*, yaitu pemberian label gramatikal pada setiap kata dalam kalimat bahasa Arab. Representasi kontekstual yang dihasilkan oleh AraBERT memungkinkan model membedakan fungsi kata yang sama secara bentuk tetapi berbeda secara gramatikal bergantung pada konteksnya. Oleh karena itu, penggunaan AraBERT dipandang relevan dan tepat untuk analisis struktur gramatikal bahasa Arab berbasis *token classification*.

2.2.6 Fine-Tuning Model Bahasa

Fine-tuning merupakan proses pelatihan lanjutan terhadap model bahasa pra-latih dengan menggunakan dataset yang bersifat spesifik sesuai dengan tujuan penelitian. Melalui proses ini, model tidak dilatih dari awal, melainkan menyesuaikan parameter yang telah dipelajari sebelumnya agar lebih relevan dengan tugas yang ditargetkan. Pendekatan *fine-tuning* memungkinkan

pemanfaatan pengetahuan linguistik umum yang telah dimiliki model bahasa sekaligus meningkatkan performa pada domain tertentu (Al-Hajj et al., 2021).

Dalam bidang NLP, *fine-tuning* umum diterapkan pada berbagai tugas, termasuk klasifikasi teks, *named entity recognition*, serta *sequence labeling*. Pada tugas *sequence labeling*, model dilatih untuk memberikan label pada setiap token dalam suatu urutan teks dengan mempertimbangkan konteks token tersebut dalam keseluruhan kalimat. Pendekatan ini sejalan dengan kebutuhan analisis bahasa Arab yang menekankan relasi gramatikal antar kata dalam satu struktur kalimat (Pouramini & Faili, 2024).

Fine-tuning pada penelitian ini diterapkan pada model AraBERT untuk mendukung tugas *token classification* dalam analisis kelas kata sebagai aspek gramatikal dasar bahasa Arab. Proses pelatihan diarahkan agar model mampu mengenali dan mengklasifikasikan setiap token berdasarkan informasi gramatikal yang terdapat dalam dataset anotasi, seperti jenis kata isim, fi'il, dan harf. Melalui *fine-tuning*, AraBERT tidak hanya mempelajari bentuk kata secara terpisah, tetapi juga memahami fungsi gramatikal kata berdasarkan konteksnya dalam kalimat. Model hasil *fine-tuning* ini selanjutnya dimanfaatkan sebagai komponen inti dalam sistem chatbot pembelajaran bahasa Arab untuk memberikan analisis klasifikasi kelas kata secara otomatis dan kontekstual.

2.2.7 *Hyperparameter* Pelatihan Model Transformer

Hyperparameter merupakan parameter pelatihan yang ditentukan sebelum proses pembelajaran model dimulai dan tidak diperbarui secara langsung selama pelatihan. *Hyperparameter* berperan dalam mengatur bagaimana model belajar dari

data, sehingga pemilihannya memengaruhi stabilitas pelatihan, kecepatan konvergensi, serta kemampuan generalisasi model (Vaswani et al., 2017).

Dalam pelatihan model bahasa berbasis Transformer, *hyperparameter* berfungsi sebagai faktor pengendali proses optimisasi agar model dapat menyesuaikan parameter secara efektif terhadap tugas yang ditargetkan. Oleh karena itu, pemahaman terhadap peran *hyperparameter* menjadi penting dalam perancangan eksperimen pelatihan model (Puvis De Chavannes et al., 2021).

2.2.8 Perbedaan Domain Qur'anic dan MSA

Bahasa Arab Al-Qur'an (Qur'anic Arabic) memiliki karakteristik linguistik yang berbeda dengan *Modern Standard Arabic* (MSA) dalam aspek kosakata, morfologi, dan struktur sintaksis. Dataset MASAQ yang digunakan dalam penelitian ini berasal dari teks Al-Qur'an, sedangkan implementasi chatbot diarahkan pada kalimat bahasa Arab umum (MSA). Perbedaan domain ini berpotensi menimbulkan fenomena domain mismatch, yaitu kondisi ketika distribusi data pelatihan dan data penggunaan berbeda sehingga memengaruhi performa model.

Dalam tugas *token classification*, perbedaan antara teks Qur'anic dan MSA dapat memengaruhi kinerja model, terutama ketika model dihadapkan pada pola kalimat yang tidak banyak muncul saat pelatihan. Walaupun pembagian kelas kata tetap sama, variasi penggunaan kosakata dan struktur kalimat pada MSA dapat menyebabkan perbedaan tingkat akurasi. Oleh karena itu, perbedaan domain ini dipahami sebagai keterbatasan penelitian dan menjadi peluang untuk pengembangan selanjutnya.

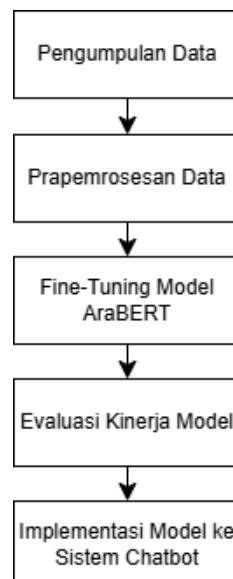
BAB III

METODOLOGI PENELITIAN

3.1 Desain Penelitian

Desain penelitian merupakan rancangan sistematis yang digunakan untuk memandu pelaksanaan penelitian agar mampu menjawab tujuan yang telah ditetapkan (Indu & K, 2020). Penelitian ini menggunakan pendekatan penelitian dan pengembangan (*research and development*) yang bertujuan menghasilkan sistem chatbot pembelajaran bahasa Arab berbasis *Natural Language Processing* (NLP) melalui penerapan fine-tuning model bahasa pralatih AraBERT.

Tahapan penelitian meliputi pengumpulan dan prapemrosesan data, *fine-tuning* model AraBERT, evaluasi kinerja model, serta implementasi model ke dalam sistem chatbot pembelajaran. Alur desain penelitian secara keseluruhan ditunjukkan pada Gambar 3.1.



Gambar 3.1 Desain Penelitian

3.2 Data Penelitian

Data penelitian merupakan omponen utama dalam proses pelatihan model *Natural Language Processing* berbasis *supervised learning* (Badawi, 2021). Pada penelitian ini, data digunakan sebagai sumber pembelajaran model AraBERT untuk melakukan klasifikasi jenis kata bahasa Arab pada tingkat token (*token classification*).

3.2.1 Sumber Data

Data yang digunakan dalam penelitian ini berasal dari dataset publik *Morphologically and Syntactically Annotated Quran Dataset* (MASAQ), yaitu corpus linguistik bahasa Arab yang menyediakan anotasi morfologi dan sintaksis ayat-ayat Al-Qur'an hingga tingkat kata. Dataset ini diperoleh melalui repositori ilmiah yang terindeks dan tersedia secara terbuka melalui platform Medeley Data, serta telah digunakan dalam berbagai penelitian sebelumnya di bidang linguistik Arab dan *Natural Language Processing* (NLP) (Sawalha, 2024).

Dalam penelitian ini, dataset MASAQ digunakan sebagai corpus linguistik untuk keperluan analisis struktur gramatikal bahasa Arab, bukan untuk tujuan tafsir atau penafsiran makna ayat Al-Qur'an. Pemanfaatan dataset publik ini bertujuan untuk menjamin kualitas anotasi, konsistensi data, serta memungkinkan replikasi penelitian oleh peneliti lain. Dataset MASAQ pada sumber aslinya tersedia dalam format *Comma-Separated Values* (CSV), yang merepresentasikan setiap unit kata sebagai satu baris data lengkap dengan atribut linguistiknya.

3.2.2 Struktur Dataset MASAQ

Dataset MASAQ memiliki struktur data yang terorganisasi berdasarkan ayat, di mana setiap ayat diuraikan hingga tingkat kata. Setiap kata dilengkapi dengan berbagai atribut linguistik yang merepresentasikan karakteristik gramatikal bahasa Arab (Al-Sheikh & Mohd, 2019). Atribut utama yang dimanfaatkan dalam penelitian ini meliputi:

- a. Word, yaitu bentuk kata berharakat dalam teks Arab
- b. Morph_type, yang menunjukkan jenis morfologi kata (misalnya *Stem*, *Prefix*, atau *Suffix*)

Penelitian ini secara khusus memfokuskan pemrosesan data pada kata dengan *Morph_type = Stem*, karena bentuk stem merepresentasikan kata utama yang memiliki fungsi sintaksis eksplisit dalam analisis klasifikasi kelas kata. Dengan demikian, token-token berupa prefiks atau sufiks tidak disertakan dalam proses pelatihan model.

Karakteristik data yang telah terannotasi secara lengkap dan konsisten ini menjadikan dataset MASAQ sesuai untuk pendekatan supervised learning, khususnya untuk tugas *sequence labeling* pada pemrosesan bahasa alami bahasa Arab.

3.3 Preprocessing Dataset

Preprocessing data dilakukan untuk menyiapkan dataset agar sesuai dengan kebutuhan pemodelan *sequence labeling* berbasis Transformer serta mendukung

proses pelatihan model AraBERT (Matrane et al., 2024). Dataset inguistik bahasa Arab pada penelitian ini didefinisikan sebagai:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \quad (3.1)$$

dengan x_i merupakan urutan token kata dalam satu ayat atau kalimat bahasa Arab, y_i adalah urutan label kelas kata yang bersesuaian, dan N adalah jumlah sampel data.

3.3.1 Penyaringan *Morph_type*

Tahap awal *preprocessing* dilakukan dengan penyaringan data berdasarkan atribut *Morph_type*, sehingga hanya token dengan nilai *Stem* yang dipertahankan. Langkah ini dilakukan karena bentuk *stem* merepresentasikan kata inti yang memiliki fungsi sintaksis eksplisit dalam analisis klasifikasi kelas kata. Selanjutnya, data diurutkan berdasarkan nomor surah, nomor ayat, dan urutan kata untuk menjaga konsistensi struktur token dalam setiap ayat.

3.3.2 Normalisasi Teks Arab

Tahap normalisasi teks dilakukan menggunakan modul *ArabertPreprocessor*. Normalisasi ini bertujuan menyesuaikan representasi teks Arab agar selaras dengan format yang digunakan pada proses *pretraining* model AraBERT (Antoun et al., 2020).

Proses normalisasi meliputi penghilangan tanda diakritik (harakat), standarisasi bentuk huruf arab, serta pembersihan karakter yang tidak relevan. Sebagai contoh, kata berharakat seperti “بِسْمِ” dinormalisasi menjadi “بسم”. Tahap ini

penting untuk menjaga konsistensi representasi token dan meminimalkan kemungkinan token tidak dikenali oleh model.

3.3.3 Pembentukan Sequence

Setelah proses normalisasi selesai, token dikelompokkan berdasarkan nomor surah dan nomor ayat sehingga setiap ayat direpresentasikan sebagai satu *sequence* token yang utuh.

Setiap *sequence* token kemudian dipasangkan dengan *sequence label* kelas kata yang bersesuaian sehingga membentuk pasangan data X_i, Y_i untuk proses *supervised learning*. Label kelas kata didefinisikan sebagai kelas diskrit:

$$Y_i \in \{1, 2, \dots, C\} \quad (3.2)$$

Dengan C menyatakan jumlah kelas kata yaitu *isim*, *fi'il*, dan *harf*. Penyederhanaan label dari anotasi asli dataset dilakukan untuk menyesuaikan kebutuhan sistem chatbot pembelajaran bahasa Arab.

3.3.4 Tokenisasi dan Label Alignment

Token hasil preprocessing selanjutnya ditokenisasi menggunakan tokenizer bawaan AraBERT. Proses tokenisasi menggunakan mekanisme *WordPiece* yang memungkinkan satu kata dipecah menjadi beberapa *subword token* sesuai kosakata model.

Karena tugas yang digunakan adalah *sequence labeling*, maka dilakukan proses *label alignment* untuk memastikan bahwa setiap *subword* hasil tokenisasi tetap memiliki label kelas kata yang sesuai dengan token asalnya

Sebagai contoh, ayat: " إِنَّ اللَّهَ يَغْفِرُ الذُّنُوبَ " direpresentasikan pada level kata sebagai:

Tokens : ["إِنَّ", "اللَّهِ", "يَغْفِرُ", "الذُّنُوبَ"]

Labels : ["HARF", "ISIM", "FI'IL", "ISIM"]

Dalam konteks penelitian ini, kelas *Isim* dapat dipahami sebagai nomina, *Fi'il* sebagai verba, dan *Harf* sebagai partikel.

3.3.5 Representasi Input Model

Hasil tokenisasi kemudian dikonversi ke bentuk representasi numerik berupa *input_ids* dan *attention_mask* untuk digunakan sebagai masukan model AraBERT.

Attention mask digunakan untuk membedakan token valid dan token *padding* sehingga model hanya memproses token yang relevan pada setiap *sequence*. Representasi input ini menjadikan dataset kompatibel dengan arsitektur Transformer yang digunakan pada proses *fine-tuning* model AraBERT.

3.3.6 Karakteristik Data setelah *Preprocessing*

Setelah melalui tahapan preprocessing, dataset akhir yang digunakan dalam penelitian ini terdiri dari 6.217 *sequence* ayat bahasa Arab yang direpresentasikan sebagai urutan token. Total data yang digunakan dalam proses pelatihan berjumlah 77.368 token.

Distribusi kelas kata dianalisis pada tingkat token karena tugas yang dilakukan adalah *token classification*. Hasil analisis distribusi menunjukkan bahwa kelas *isim* merupakan kelas dominan dengan 39.098 token, diikuti oleh kelas *fi'il*

sebanyak 19.345 token, dan kelas *harf* sebanyak 18.925 token. Distribusi data ditunjukkan pada Tabel 3.1

Tabel 3.1 Distribusi Token per Kelas

Kelas	Jumlah Token	Persentase (%)
ISIM	39.098	50.54
FI'IL	19.345	25.00
HARF	18.925	24.46
Total	77.368	100

Distribusi data menunjukkan bahwa kelas *isim* memiliki proporsi paling besar dibandingkan kelas lainnya. Kondisi ini mencerminkan karakteristik umum struktur bahasa Arab, khususnya pada corpus Al-Qur'an, yang banyak mengandung bentuk nomina dalam susunan kalimatnya. Meskipun demikian, ketiga kelas kata tetap memiliki representasi data yang memadai untuk digunakan dalam proses pelatihan model klasifikasi token berbasis *supervised learning*. Dataset hasil *preprocessing* tersebut selanjutnya digunakan pada tahap *fine-tuning* model AraBERT untuk melakukan klasifikasi kelas kata bahasa Arab pada tingkat token.

3.4 Fine-Tuning Model AraBERT

Penelitian ini menggunakan pendekatan ini menggunakan pendekatan *supervised learning* dengan melakukan *fine-tuning* pada model AraBERT, yaitu model bahasa Arab berbasis arsitektur Transformer yang telah dilatih sebelumnya (*pretrained*) (El Moubtahij et al., 2022). *Fine-tuning* dilakukan untuk tugas *token classification (sequence labeling)*, yaitu mengklasifikasikan setiap token ke dalam kelas kata *isim*, *fi'il*, atau *harf*.

3.4.1 Arsitektur Model AraBERT

Pada penelitian ini, AraBERT berfungsi sebagai *encoder* yang menghasilkan representasi kontekstual untuk setiap token dalam *sequence*. Representasi tersebut kemudian dipetakan ke ruang kelas menggunakan *linear classification layer*. Secara matematis, proses forward pass dapat diringkas sebagai:

$$H = \text{AraBERT}_0(X) \quad (3.3)$$

Dengan $H \in \mathbb{R}^{T \times d}$, yaitu T adalah panjang sequence token dan d adalah dimensi *embedding*. Hasil representasi token kemudian digunakan untuk memprediksi kelas kata pada setiap token.

3.4.2 Fungsi Loss

Proses pelatihan model menggunakan *Cross-Entropy Loss* pada tingkat token. Fungsi ini mengukur perbedaan antara distribusi probabilitas prediksi model dan label kelas kata sebenarnya. *Loss* dihitung untuk setiap token dalam satu batch dan dirata-ratakan sebagai objective function yang diminimalkan selama pelatihan (Antoun et al., 2020).

Penggunaan *loss* berbasis token memastikan bahwa model belajar mengoptimalkan prediksi untuk setiap kata dalam *sequence* dengan tetap mempertimbangkan konteks antar token.

3.4.3 Optimasi parameter

Optimasi parameter dilakukan menggunakan algoritma AdamW. AdamW merupakan varian dari Adam yang menerapkan mekanisme *weight decay* secara

terpisah untuk meningkatkan stabilitas pelatihan dan mengurangi risiko *overfitting* pada model berbasis Transformer. Pembaruan parameter dilakukan berdasarkan *gradien loss* terhadap parameter model dengan penyesuaian *learning rate* adaptif.

3.4.4 Konfigurasi *Hyperparameter* Pelatihan

Proses *fine-tuning* model AraBERT dilakukan menggunakan sejumlah konfigurasi *hyperparameter* pelatihan yang telah disesuaikan dengan kebutuhan tugas *token classification* (Puvis De Chavannes et al., 2021). Penentuan konfigurasi dilakukan untuk menjaga kestabilan proses pelatihan sekaligus menyesuaikan dengan keterbatasan sumber daya komputasi yang digunakan selama penelitian.

Konfigurasi yang digunakan meliputi *learning rate*, *batch size*, jumlah *epoch* maksimum, penggunaan *optimizer* AdamW, serta mekanisme *early stopping*. Selain itu, proses evaluasi model dilakukan secara berkala pada setiap epoch menggunakan data validasi untuk memantau performa pelatihan model.

Pemilihan konfigurasi tersebut bertujuan untuk memperoleh keseimbangan antara stabilitas pelatihan, kemampuan generalisasi model, dan efisiensi proses komputasi. Seluruh konfigurasi pelatihan diterapkan secara konsisten pada setiap skenario eksperimen agar perbandingan performa model dapat dilakukan secara objektif.

3.4.5 Algoritma *Fine-Tuning* AraBERT

Setelah arsitektur model, fungsi *loss*, metode optimasi, dan konfigurasi *hyperparameter* dijelaskan, tahap selanjutnya adalah merumuskan proses *fine-tuning* AraBERT secara operasional. *Fine-tuning* pada penelitian ini menggunakan

pendekatan *supervised learning*, yaitu setiap token dalam dataset berlabel digunakan sebagai target prediksi untuk menyesuaikan parameter model AraBERT *pretrained* pada tugas token classification.

Proses ini bertujuan agar model mampu mengklasifikasikan jenis kata bahasa Arab (*isim*, *fi'il*, dan *harf*) berdasarkan konteks token dalam kalimat. Secara umum, proses ini menerima masukan berupa dataset latih berlabel, model AraBERT *pretrained*, serta parameter pelatihan, dan menghasilkan model AraBERT terlatih yang digunakan pada tahap inferensi (Pouramini & Faili, 2024).

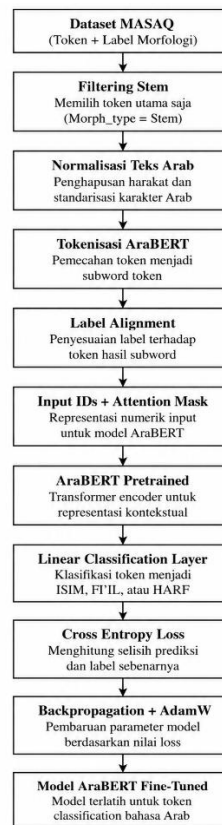
Spesifikasi input dan output proses *fine-tuning* disajikan pada Tabel 3.2.

Tabel 3.2 Input dan Output Proses *Fine-Tuning*

Input	Output
Dataset latih D_{train} , model AraBERT pretrained, learning rate α , jumlah epoch E , dan ukuran batch B	Model AraBERT terlatih

Setelah spesifikasi input dan output ditetapkan, penelitian ini merancang alur proses *fine-tuning* model AraBERT untuk melakukan klasifikasi kelas kata bahasa Arab. Alur tersebut mencakup tahapan preprocessing dataset, tokenisasi teks, penyesuaian label terhadap token hasil tokenisasi, pembentukan representasi input model, hingga proses pelatihan dan optimasi model menggunakan AraBERT pretrained. Tahapan *fine-tuning* dimulai dari dataset MASAQ berlabel morfologi, penyaringan token berdasarkan *Morph_type*, normalisasi teks Arab, tokenisasi menggunakan WordPiece tokenizer, proses label alignment, pembentukan representasi input model berupa *input_ids* dan *attention_mask*, hingga proses pelatihan menggunakan model AraBERT pretrained. Selanjutnya, hasil representasi kontekstual token diproses menggunakan linear classification layer untuk melakukan klasifikasi kelas kata. Proses pelatihan dioptimalkan

menggunakan Cross-Entropy Loss dan optimizer AdamW hingga diperoleh model AraBERT hasil fine-tuning. Alur proses fine-tuning model AraBERT pada penelitian ini ditunjukkan pada Gambar 3.2.



Gambar 3.2 Pipeline *Fine-Tuning* Model AraBERT

Berdasarkan rancangan tersebut, algoritma *fine-tuning* AraBERT dirumuskan secara operasional sebagai berikut.

Algoritma 1 *Fine-Tuning* AraBERT

Input: Dataset latih berlabel D_{train} , model AraBERT *pretrained*, learning rate α , jumlah *epoch* E , dan ukuran *batch* B

Output: Model AraBERT hasil *fine-tuning*

1. Memuat model AraBERT *pretrained*.
2. Menyesuaikan *output layer* untuk tugas *token classification*.

3. Menginisialisasi optimizer AdamW dengan parameter pelatihan yang telah ditentukan.
4. Membagi dataset latih ke dalam *batch* berukuran B.
5. Untuk setiap *epoch* ke-1 hingga *E*, dilakukan langkah berikut:
 - a. Mengambil satu *batch* data latih.
 - b. Melakukan *forward pass* untuk memperoleh representasi kontekstual token.
 - c. Menghitung nilai *cross-entropy loss* antara prediksi model dan label sebenarnya.
 - d. Melakukan *backpropagation* untuk menghitung *gradien* parameter model.
 - e. Memperbarui parameter model menggunakan optimizer AdamW.
6. Mengulangi proses hingga seluruh *epoch* selesai.
7. Menyimpan model AraBERT hasil *fine-tuning* untuk digunakan pada tahap inferensi.

Algoritma ini memastikan bahwa model AraBERT tidak hanya memanfaatkan pengetahuan linguistik umum hasil *pretraining*, tetapi juga mampu beradaptasi secara spesifik terhadap tugas klasifikasi jenis kata bahasa Arab berdasarkan dataset anotasi morfologi. Model hasil *fine-tuning* inilah yang selanjutnya digunakan sebagai komponen inti dalam implementasi sistem chatbot pembelajaran bahasa Arab.

3.5 Evaluasi Model

Tahap evaluasi model dilakukan untuk mengukur kemampuan model AraBERT hasil *fine-tuning* dalam melakukan klasifikasi kelas kata bahasa Arab.

3.5.1 Metrik Evaluasi Model

Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang umum digunakan pada tugas klasifikasi NLP (Yacouby & Axman, 2020).

Akurasi merupakan metrik yang mengukur proporsi token yang diklasifikasikan dengan benar terhadap seluruh token pada data uji. Perhitungan *accuracy* dirumuskan pada persamaan 3.4 berikut ini:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4)$$

Precision menunjukkan tingkat ketepatan model dalam memprediksi suatu kelas kata tertentu. Rumus *precision* ditunjukkan pada persamaan 3.5 berikut ini:

$$Precision = \frac{TP}{TP + FP} \quad (3.5)$$

Recall mengukur kemampuan model dalam mengenali seluruh token yang benar-benar termasuk dalam suatu kelas. Perhitungan *recall* dirumuskan pada persamaan 3.6 berikut ini:

$$Recall = \frac{TP}{TP + FN} \quad (3.6)$$

F1-Score merupakan rata-rata harmonik antara *precision* dan *recall* yang digunakan untuk menyeimbangkan kedua metrik tersebut. Perhitungan *F1-score* dirumuskan pada persamaan 3.7 berikut ini:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (3.7)$$

3.5.2 Analisis Generalisasi Domain

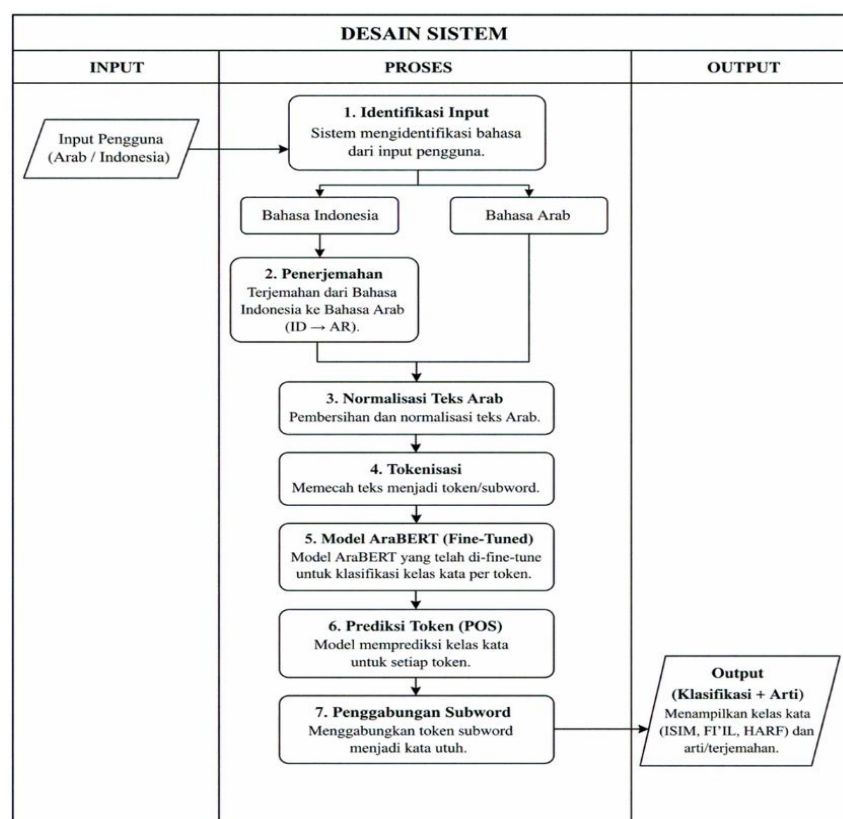
Dataset pelatihan dalam penelitian ini berasal dari corpus *Qur'anic Arabic*, sedangkan pengujian tambahan dilakukan pada teks *Modern Standard Arabic* (MSA). Perbedaan karakteristik linguistik antara kedua domain tersebut berpotensi menimbulkan fenomena *domain shift*, yaitu kondisi ketika distribusi data pelatihan berbeda dari distribusi data pengujian. *Qur'anic Arabic* merupakan bentuk bahasa Arab klasik dengan karakteristik leksikal dan sintaktis yang khas, seperti penggunaan kosakata arkais dan struktur morfologi yang lebih kompleks. Sebaliknya, MSA merupakan bahasa Arab modern yang umum digunakan dalam konteks pendidikan, jurnalistik, dan komunikasi formal dengan struktur kalimat yang relatif lebih sederhana.

Perbedaan kosakata, variasi morfologi, dan pola sintaktis antara *Qur'anic Arabic* dan MSA dapat mempengaruhi representasi kontekstual token yang dihasilkan oleh model berbasis Transformer. Model yang dilatih pada domain *Qur'anic Arabic* cenderung mempelajari distribusi konteks yang spesifik terhadap gaya bahasa klasik sehingga performa model berpotensi menurun ketika diuji pada domain MSA. Oleh karena itu, evaluasi pada data MSA dalam penelitian ini dianalisis sebagai pengujian generalisasi lintas domain (*cross-domain generalization*) untuk mengukur kemampuan model dalam mempertahankan akurasi klasifikasi jenis kata pada domain bahasa yang berbeda.

3.6 Desain Sistem Chatbot Pembelajaran

Salah satu tahapan penting dalam penelitian ini adalah perancangan desain sistem chatbot pembelajaran bahasa Arab. Desain sistem dirancang untuk mengintegrasikan model AraBERT hasil *fine-tuning* ke dalam proses analisis kelas kata bahasa Arab secara otomatis melalui antarmuka sistem (Mulyadi, 2013).

Representasi desain sistem chatbot ditunjukkan pada Gambar 3.3.



Gambar 3.3 Desain Sistem Chatbot

Gambar 3.3 menggambarkan alur implementasi model AraBERT hasil fine-tuning ke dalam sistem chatbot pembelajaran bahasa Arab. Sistem menerima input berupa teks bahasa Arab maupun bahasa Indonesia. Apabila input diberikan dalam bahasa Indonesia, sistem terlebih dahulu melakukan proses penerjemahan ke

bahasa Arab sebelum memasuki tahap pemrosesan utama. Selanjutnya, teks diproses melalui tahap normalisasi dan tokenisasi sebelum dilakukan klasifikasi token menggunakan model AraBERT hasil fine-tuning.

Pada tahap inferensi, model melakukan proses *sequence labeling* atau *token classification* untuk memprediksi kelas kata pada setiap token dalam kalimat. Hasil prediksi kemudian diproses kembali melalui tahap *post-processing* berupa penggabungan subword token sehingga output yang ditampilkan menjadi lebih mudah dipahami pengguna. Sistem kemudian menyajikan hasil analisis kelas kata dalam bentuk yang lebih sederhana dan edukatif untuk mendukung proses pembelajaran bahasa Arab dasar.

Meskipun sistem dirancang untuk menerima input bahasa Indonesia, evaluasi utama model tetap dilakukan menggunakan teks bahasa Arab secara langsung untuk menjaga konsistensi dengan data pelatihan serta memastikan bahwa performa yang diukur merefleksikan kemampuan inti model tanpa dipengaruhi oleh kemungkinan kesalahan translasi.

Secara operasional, alur inferensi chatbot pembelajaran bahasa Arab pada penelitian ini dirumuskan dalam bentuk algoritma sebagai berikut.

Algoritma 2 Inferensi Chatbot Pembelajaran Bahasa Arab

Input: Teks dari pengguna

Output: Respon chatbot berupa analisis jenis kata

1. Terima teks/kalimat dari pengguna.

2. Identifikasi bahasa input. Jika input berupa bahasa Indonesia, lakukan penerjemahan ke bahasa Arab, sedangkan jika input berupa bahasa Arab, proses dilanjutkan ke tahap berikutnya.
3. Lakukan *preprocessing* ringan dan tokenisasi teks bahasa Arab menggunakan tokenizer AraBERT.
4. Bentuk pasangan input (*input_ids*, *attention_mask*).
5. Lakukan inferensi *token classification* menggunakan model AraBERT terlatih.
6. Peroleh prediksi label kelas kata untuk setiap token.
7. Lakukan *post-processing* dengan memetakan label ke penjelasan linguistik.
8. Susun respon chatbot berdasarkan hasil analisis token.
9. Tampilkan respon chatbot kepada pengguna.

Algoritma tersebut menggambarkan alur utama inferensi pada sistem chatbot, mulai dari penerimaan input pengguna hingga penyajian hasil klasifikasi kelas kata. Untuk memperjelas implementasi setiap tahapan pada sistem chatbot, rincian proses inferensi dijelaskan lebih lanjut pada subbab berikut.

3.6.1 Input Pengguna

Sistem chatbot dirancang untuk menerima masukan berupa teks bahasa Arab maupun bahasa Indonesia dalam bentuk kalimat atau frasa sederhana. Input diberikan langsung oleh pengguna melalui antarmuka chatbot dan diproses sebagai data teks berbentuk string. Kemampuan menerima dua jenis bahasa ini bertujuan untuk meningkatkan fleksibilitas penggunaan sistem, khususnya bagi pengguna pemula yang belum sepenuhnya mampu menulis kalimat dalam bahasa Arab.

Sebagai contoh, pengguna dapat memberikan input dalam bahasa Arab seperti “الطالب يكتب الدرس” maupun dalam bahasa Indonesia seperti “siswa menulis pelajaran”. Meskipun sistem menyediakan dukungan input bahasa Indonesia, proses inti klasifikasi tetap dilakukan terhadap representasi teks bahasa Arab agar sesuai dengan karakteristik model AraBERT yang digunakan dalam penelitian ini.

3.6.2 Identifikasi Bahasa Input

Setelah menerima input dari pengguna, sistem melakukan proses identifikasi bahasa untuk menentukan jalur pemrosesan berikutnya. Identifikasi dilakukan secara *rule-based* dengan memeriksa keberadaan karakter *Unicode* Arab pada teks masukan (Alruqi & Alzahrani, 2023). Jika sistem mendeteksi adanya karakter Arab, maka input diproses langsung ke tahap normalisasi teks Arab. Sebaliknya, apabila input tidak mengandung karakter Arab, sistem menganggap input sebagai bahasa Indonesia dan meneruskannya ke tahap penerjemahan.

Pendekatan *rule-based* digunakan karena memiliki implementasi yang ringan dan cukup efektif untuk membedakan teks Arab dan non-Arab pada sistem chatbot pembelajaran ini. Tahap identifikasi bahasa penting untuk memastikan bahwa seluruh input yang diproses oleh model AraBERT berada dalam representasi bahasa yang sesuai.

3.6.3 Penerjemahan Bahasa Indonesia ke Bahasa Arab

Apabila input yang diberikan pengguna berupa bahasa Indonesia, sistem akan melakukan proses penerjemahan otomatis ke dalam bahasa Arab sebelum memasuki tahap pemrosesan utama. Proses penerjemahan dilakukan menggunakan

library *deep_translator* dengan layanan GoogleTranslator sebagai mesin translasi (Zeroual et al., 2017).

Tahap ini bertujuan untuk memungkinkan pengguna yang belum mahir menulis bahasa Arab tetap dapat menggunakan sistem chatbot pembelajaran secara interaktif. Sebagai contoh, kalimat “saya belajar bahasa Arab” diterjemahkan menjadi “أنا أتعلم اللغة العربية” sebelum diproses lebih lanjut oleh model.

Pada penelitian ini, proses penerjemahan hanya digunakan pada implementasi chatbot sebagai fitur tambahan antarmuka pengguna dan tidak digunakan dalam proses pelatihan maupun evaluasi utama model. Evaluasi model tetap dilakukan menggunakan teks bahasa Arab secara langsung agar performa yang diukur merefleksikan kemampuan klasifikasi inti model tanpa dipengaruhi kemungkinan kesalahan translasi.

3.6.4 Normalisasi Teks Arab

Teks bahasa Arab hasil input langsung maupun hasil penerjemahan selanjutnya diproses melalui tahap normalisasi teks. Tahap ini bertujuan untuk menyesuaikan representasi teks dengan format yang digunakan pada proses pretraining AraBERT sehingga token yang dihasilkan lebih konsisten dan mudah dikenali oleh model (Antoun et al., 2020).

Normalisasi dilakukan menggunakan modul *ArabertPreprocessor* yang menyediakan berbagai fungsi pembersihan teks Arab. Proses normalisasi meliputi penghilangan tanda diakritik (harakat), standarisasi bentuk huruf Arab, serta

penghapusan karakter yang tidak relevan. Sebagai contoh, token “بِسْمِ” dinormalisasi menjadi “بسم”.

Tahap normalisasi penting dalam pemrosesan bahasa Arab karena variasi penulisan huruf dan penggunaan harakat dapat mempengaruhi hasil tokenisasi serta representasi kontekstual yang dipelajari model Transformer.

3.6.5 Tokenisasi Teks

Sistem Setelah proses normalisasi selesai dilakukan, teks diproses menggunakan *tokenizer* bawaan AraBERT. Tokenisasi bertujuan untuk mengubah teks menjadi representasi token numerik yang dapat dipahami oleh model Transformer (Rogers et al., 2020).

Tokenizer AraBERT menggunakan mekanisme *WordPiece tokenization*, yaitu metode tokenisasi yang memungkinkan satu kata dipecah menjadi beberapa subword token berdasarkan kosakata model. Pendekatan ini membantu model menangani kata yang jarang muncul atau memiliki variasi morfologi kompleks dalam bahasa Arab.

Hasil tokenisasi kemudian direpresentasikan dalam bentuk pasangan *input_ids* dan *attention_mask*. *Input_ids* merepresentasikan indeks numerik token dalam vocabulary model, sedangkan *attention_mask* digunakan untuk membedakan token valid dan token padding pada setiap *sequence*.

3.6.6 Model AraBERT *Fine-Tuned*

Tahap selanjutnya adalah integrasi model AraBERT hasil *fine-tuning* ke dalam sistem chatbot pembelajaran. Model yang digunakan pada tahap ini merupakan model hasil pelatihan pada Subbab 3.5 yang telah disesuaikan untuk tugas *token classification* bahasa Arab menggunakan dataset berlabel.

Pada implementasi chatbot, model AraBERT berfungsi sebagai mesin analisis linguistik yang menerima representasi token hasil preprocessing dan menghasilkan prediksi kelas kata untuk setiap token dalam kalimat. Integrasi model dilakukan dengan memuat parameter model hasil *fine-tuning* ke dalam sistem inferensi chatbot sehingga proses klasifikasi dapat dilakukan secara otomatis terhadap input pengguna.

Dengan integrasi tersebut, sistem chatbot mampu memanfaatkan representasi kontekstual yang dipelajari model selama proses pelatihan untuk melakukan analisis jenis kata bahasa Arab secara interaktif dan *real-time*.

3.6.7 Prediksi Kelas Kata (*POS Tagging*)

Berdasarkan representasi kontekstual yang dihasilkan model, sistem melakukan prediksi kelas kata untuk setiap token dalam sequence. Proses ini termasuk ke dalam tugas *sequence labeling* atau *token classification*, yaitu pemberian label linguistik pada setiap token dalam kalimat.

Prediksi dilakukan dengan mempertimbangkan konteks antar token sehingga model tidak hanya memahami token secara individual, tetapi juga hubungan sintaktis dalam kalimat. Pada tahap ini, model menghasilkan label kelas kata berupa *isim*, *fi'il*, atau *harf* untuk setiap token yang diproses. Dalam konteks penelitian ini,

kelas *isim* merepresentasikan nomina, kelas *fi'il* merepresentasikan verba, dan kelas *harf* merepresentasikan partikel dalam tata bahasa Arab

3.6.8 Penggabungan *Subword*

Karena proses tokenisasi menggunakan mekanisme *WordPiece*, beberapa kata dapat dipecah menjadi lebih dari satu *subword* token. Oleh karena itu, sistem melakukan tahap *post-processing* untuk menggabungkan kembali *subword* menjadi bentuk kata utuh agar hasil klasifikasi lebih mudah dipahami pengguna.

Sebagai contoh, token hasil tokenisasi seperti “ال + طالب” akan digabung kembali menjadi “الطالب”. Proses penggabungan ini dilakukan setelah tahap prediksi kelas kata selesai dan sebelum hasil ditampilkan pada antarmuka chatbot.

Tahap penggabungan *subword* penting untuk meningkatkan keterbacaan output sistem serta menjaga kesesuaian hasil analisis linguistik dengan bentuk kata asli dalam bahasa Arab. Dengan demikian, pengguna dapat melihat hasil klasifikasi dalam bentuk kata yang utuh dan lebih natural.

3.6.9 Output Sistem Chatbot Pembelajaran

Tahap akhir pada sistem chatbot adalah penyajian hasil analisis kepada pengguna. Sistem menampilkan hasil klasifikasi berupa kata bahasa Arab, kelas kata hasil prediksi, serta arti atau terjemahan sederhana untuk membantu proses pembelajaran bahasa Arab (Faroid et al., 2025).

Output disajikan dalam bentuk analisis linguistik sederhana sehingga pengguna dapat memahami fungsi gramatikal setiap kata dalam kalimat secara lebih mudah. Dengan pendekatan ini, chatbot tidak hanya berfungsi sebagai sistem

klasifikasi otomatis, tetapi juga sebagai media pembelajaran interaktif dalam memahami dasar tata bahasa Arab.

Sebagai contoh, ketika pengguna memasukkan kalimat “إِنَّ الطَّالِبَ يَكْتُبُ الدَّرْسَ” sistem dapat menampilkan hasil klasifikasi kelas kata sebagaimana ditunjukkan pada Tabel 3.3.

Tabel 3.3 Contoh Hasil Prediksi Kelas Kata

Token	Prediksi
إِنَّ	<i>Harf</i>
الطَّالِبَ	<i>Isim</i>
يَكْتُبُ	<i>Fi'il</i>
الدَّرْسَ	<i>Isim</i>

Hasil tersebut menunjukkan bahwa sistem mampu mengidentifikasi kelas kata pada setiap token dalam kalimat bahasa Arab secara otomatis. Informasi ini kemudian disajikan kepada pengguna dalam bentuk yang lebih mudah dipahami sehingga dapat membantu proses pembelajaran nahwu dasar secara interaktif.

3.7 Evaluasi Sistem Chatbot Pembelajaran

Evaluasi sistem chatbot pembelajaran bahasa Arab dilakukan untuk memastikan bahwa model klasifikasi yang telah dilatih dapat berfungsi dengan baik setelah diintegrasikan ke dalam aplikasi chatbot. Evaluasi ini tidak bertujuan mengukur performa model secara statistik, melainkan menguji konsistensi perilaku sistem dalam skenario penggunaan nyata.

Pengujian dilakukan menggunakan pendekatan skenario uji dengan memberikan sejumlah kalimat sebagai input ke dalam chatbot. Sistem kemudian memproses kalimat melalui tahapan deteksi bahasa, penerjemahan (jika

diperlukan), tokenisasi, hingga klasifikasi kelas kata. Hasil prediksi yang dihasilkan diamati untuk melihat kesesuaian antara label yang diberikan sistem dengan kaidah tata bahasa Arab.

Evaluasi utama dilakukan menggunakan kalimat bahasa Arab baku (*Modern Standard Arabic/MSA*). Penggunaan kalimat bahasa Arab secara langsung bertujuan memastikan bahwa model tetap mampu melakukan klasifikasi secara konsisten setelah diintegrasikan ke dalam sistem chatbot. Pengujian ini juga memungkinkan pengamatan terhadap perilaku model pada konteks kalimat utuh, bukan hanya pada data uji terpisah.

Selain itu, dilakukan pengujian terbatas menggunakan input bahasa Indonesia untuk merepresentasikan alur penggunaan sistem secara menyeluruh (*end-to-end*). Pada skenario ini, sistem terlebih dahulu menerjemahkan kalimat ke bahasa Arab sebelum dilakukan proses klasifikasi. Pengujian ini tidak digunakan sebagai dasar evaluasi performa model, tetapi untuk memastikan bahwa seluruh komponen sistem dapat bekerja secara terintegrasi

Hasil evaluasi disajikan dalam bentuk tabel perbandingan antar token, label aktual, dan label prediksi. Pendekatan ini digunakan untuk mengidentifikasi pola kesalahan dan memastikan chatbot mampu memberikan analisis kelas kata secara konsisten pada berbagai struktur kalimat.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Data Penelitian

Dataset yang digunakan dalam penelitian ini adalah *Morphologically and Syntactically Annotated Quran Dataset* (MASAQ), yaitu dataset publik yang menyediakan anotasi morfologi dan sintaksis Al-Qur'an pada tingkat token. Secara keseluruhan, dataset ini terdiri dari 157.677 unit morfologis dengan 131 label morfologi yang beragam.

Jumlah label yang besar tersebut menunjukkan kekayaan informasi linguistik dalam MASAQ. Namun, karena penelitian ini berfokus pada tugas token classification untuk mengidentifikasi kelas kata dasar bahasa Arab, maka tidak seluruh label digunakan secara langsung. Oleh karena itu, dilakukan proses label mapping untuk menyederhanakan 131 label morfologi menjadi tiga kelas utama, yaitu: *isim*, *fi'il* dan *harf*.

Penyederhanaan ini dilakukan agar sesuai dengan tujuan penelitian, yaitu membangun model yang mampu mengenali kategori kelas kata dasar dalam konteks pembelajaran bahasa Arab. Karena MASAQ berbasis segmentasi morfologis (satu kata dapat terpecah menjadi beberapa unit seperti *prefiks*, *stem*, dan *sufiks*), penelitian ini hanya menggunakan unit *stem* sebagai representasi kata utama. Unit morfologis yang tidak merepresentasikan kata mandiri atau tidak termasuk dalam tiga kelas tersebut dieliminasi.

Setelah proses penyaringan dan pemetaan label, diperoleh dataset antara (*intermediate dataset*) yang selanjutnya digunakan pada tahap preprocessing sebelum proses *fine-tuning* model.

4.2 Hasil Preprocessing

Tahap preprocessing dilakukan untuk menyesuaikan dataset hasil penyederhanaan label dengan kebutuhan model berbasis Transformer. Pada tahap ini, *token linguistik* yang telah dipetakan ke dalam tiga kelas utama masih berada pada level kata. Tabel 4.1 menunjukkan contoh bentuk dataset antara sebelum dilakukan tokenisasi *subword*.

Tabel 4.1 Contoh *Intermediate Dataset*

Sura_No	Verse_No	Token	Label
1	1	بِ	HARF
1	1	اَسْمِ	ISIM
1	1	اللّٰهِ	ISIM
1	1	الرَّحْمٰنِ	ISIM

Selanjutnya, dilakukan tokenisasi *subword* menggunakan tokenizer dari model Hugging Face untuk varian *Aubmindlab AraBERT v2*. Karena tokenisasi *subword* dapat memecah satu kata menjadi beberapa bagian, dilakukan proses *label alignment* agar setiap *subword* tetap mempertahankan label token asalnya. Token khusus (*special tokens*) tidak dilibatkan dalam perhitungan loss selama pelatihan.

Hasil akhir preprocessing adalah dataset dalam format yang kompatibel dengan pustaka Transformer pada platform Hugging Face, yang terdiri atas tiga komponen utama: *input_ids*, *attention_mask*, dan *labels* yang siap digunakan pada tahap pelatihan model. Contoh struktur data tersebut ditunjukkan pada Tabel 4. 2.

Tabel 4. 2 Struktur Data Dataset (HuggingFace)

Fitur	Contoh Nilai	Keterangan
<i>Input_ids</i>	[33, 32, 32, 32, 34, ...]	Representasi numerik token hasil tokenisasi <i>subword</i>
<i>Attention_mask</i>	[1, 1, 1, 1, 1, ...]	Penanda token valid dan padding
<i>labels</i>	[-100, 2, 0, 0, -100, ...]	Label kelas token; -100 diabaikan saat pelatihan

Nilai -100 digunakan untuk menandai token khusus hasil tokenisasi subword yang tidak dilibatkan dalam perhitungan loss selama proses pelatihan model, sehingga proses pelatihan hanya berfokus pada token linguistik yang relevan.

4.3 Data Splitting dan Skema Eksperimen

Pada tahap ini dilakukan perancangan skema eksperimen untuk mengevaluasi performa model AraBERT pada tugas *token classification* kelas kata bahasa Arab. Dataset yang telah melalui proses preprocessing, normalisasi, tokenisasi awal, dan pelabelan kelas kata selanjutnya dibagi kedalam beberapa subset data untuk kebutuhan pelatihan, validasi, dan pengujian model.

Tujuan utama dari tahap ini adalah untuk memperoleh konfigurasi pembagian data yang mampu menghasilkan performa model terbaik sekaligus tetap mempertahankan kemampuan generalisasi terhadap data baru. Oleh karena itu, penelitian ini tidak hanya menggunakan satu rasio pembagian data, tetapi melakukan beberapa skenario eksperimen rasio split data untuk dianalisis secara komparatif.

4.3.1 Pembagian Data dan *Sequence-Level Splitting*

Pembagian data dilakukan pada tingkat ayat (*sequence-level splitting*) untuk mencegah terjadinya data *leakage* (munculnya data yang sama atau sangat mirip

pada data latih dan data uji). Setiap ayat direpresentasikan sebagai satu *sequence* yang terdiri atas kumpulan token bahasa Arab beserta label kelas katanya.

Pendekatan *sequence-level splitting* digunakan karena pada tugas *token classification*, hubungan antar token dalam satu kalimat sangat memengaruhi konteks linguistik. Oleh karena itu, token dari ayat yang sama tidak boleh muncul pada data latih dan data uji secara bersamaan agar evaluasi model benar-benar mengukur kemampuan model dalam mengenali pola pada data baru.

Pembagian data dilakukan pada tingkat ayat (*sequence-level splitting*) dengan mekanisme pengacakan data untuk menghindari bias distribusi, namun tetap menjaga agar setiap ayat utuh berada pada satu subset yang sama.

Selain pembagian pada tingkat ayat, dilakukan pula analisis distribusi jumlah token pada setiap subset data. Analisis ini penting karena model Transformer bekerja pada level token, sehingga jumlah token pada data pelatihan akan memengaruhi proses pembelajaran representasi bahasa oleh model.

4.3.2 Skenario Rasio Split Data

Pada penelitian ini digunakan tiga skenario rasio split data, yaitu:

1. 70:15:15
2. 80:10:10
3. 90:5:5

Eksperimen beberapa rasio split dilakukan untuk menganalisis pengaruh proporsi data training terhadap performa hasil *fine-tuning* model AraBERT. Semakin besar proporsi data training, model memiliki peluang lebih besar untuk mempelajari pola distribusi token dan konteks linguistik bahasa Arab secara lebih

mendalam. Namun demikian, ukuran data validation dan testing yang terlalu kecil juga dapat menyebabkan evaluasi performa menjadi kurang stabil dan kurang representatif.

Pada skenario 70:15:15, sebanyak 70% data digunakan sebagai data training, sedangkan masing-masing 15% digunakan sebagai data validation dan testing. Rasio ini memberikan ukuran data evaluasi yang relatif besar sehingga hasil validasi dan pengujian menjadi lebih stabil.

Pada skenario 80:10:10, proporsi data training ditingkatkan menjadi 80%, sedangkan masing-masing 10% digunakan sebagai validation dan testing. Rasio ini dirancang untuk memberikan keseimbangan antara kemampuan pembelajaran model dan reliabilitas evaluasi.

Sementara itu, pada skenario 90:5:5, sebagian besar data digunakan sebagai training sehingga model memperoleh data pembelajaran paling banyak. Namun ukuran validation dan testing menjadi lebih kecil dibandingkan dua skenario lainnya sehingga reliabilitas evaluasi perlu dianalisis secara hati-hati. Distribusi pembagian data pada masing-masing skenario ditunjukkan pada Tabel 4.3.

Tabel 4.3 Distribusi Data Train, Validation, dan Test pada Setiap Skenario Split

Rasio Split	Train Token	Validation Token	Test Token
70:15:15	54.329	11.230	11.746
80:10:10	62.020	7.527	7.821
90:5:5	69.590	3.842	3.936

Seluruh skenario eksperimen menggunakan konfigurasi *hyperparameter* yang sama agar perbedaan performa model terutama dipengaruhi oleh perbedaan proporsi data training, validation, dan testing.

4.3.3 Tujuan dan Desain Eksperimen

Eksperimen pada penelitian ini dirancang untuk mengevaluasi kemampuan model *pretrained* AraBERT setelah dilakukan proses *fine-tuning* pada tugas klasifikasi kelas kata bahasa Arab. Fokus utama eksperimen adalah menganalisis pengaruh rasio pembagian data terhadap performa model dan kemampuan generalisasi terhadap data baru.

Seluruh eksperimen dilakukan menggunakan arsitektur dan konfigurasi pelatihan yang sama sehingga hasil perbandingan antar skenario split data dapat dianalisis secara objektif. Setiap model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, serta *validation loss* pada setiap *epoch* pelatihan.

Selain itu, penelitian ini juga menerapkan mekanisme *early stopping* untuk mencegah *overfitting* apabila performa validasi tidak lagi mengalami peningkatan. Model terbaik pada masing-masing skenario dipilih berdasarkan nilai *validation loss* terendah selama proses *fine-tuning*.

Hasil dari seluruh eksperimen selanjutnya dibandingkan untuk menentukan model terbaik yang akan digunakan pada tahap evaluasi data uji dan integrasi ke dalam sistem chatbot pembelajaran bahasa Arab.

4.4 Proses *Fine-Tuning* Model AraBERT

Pada penelitian ini, proses *fine-tuning* dilakukan menggunakan model *pretrained* AraBERT yang telah memiliki representasi bahasa Arab hasil *pretraining* pada *corpus* skala besar. Model *pretrained* tersebut kemudian disesuaikan terhadap tugas *token classification* untuk klasifikasi kelas kata bahasa Arab, yaitu *isim*, *fi'il*, dan *harf*.

Fine-tuning dilakukan bukan dengan melatih model dari awal (*training from scratch*), melainkan dengan memperbarui parameter bobot (*weights*) pada model *pretrained* agar representasi bahasa yang telah dipelajari sebelumnya dapat diadaptasikan secara khusus terhadap domain dan tugas penelitian.

Secara umum, proses *fine-tuning* meliputi tahap tokenisasi menggunakan tokenizer bawaan AraBERT, penyesuaian label token (*label alignment*), pembentukan representasi input model, pelatihan model menggunakan optimizer AdamW, evaluasi performa pada data validasi, serta pemilihan checkpoint terbaik menggunakan mekanisme *early stopping*.

4.4.1 Inisialisasi Tokenizer AraBERT

Tahap pertama pada proses *fine-tuning* adalah inisialisasi *tokenizer* bawaan AraBERT menggunakan model *aubmindlab/bert-base-arabertv02*. Tokenizer ini digunakan untuk mengubah teks bahasa Arab menjadi representasi token yang sesuai dengan *vocabulary pretrained* AraBERT.

Tokenizer AraBERT menggunakan pendekatan *subword tokenization*, yaitu proses pemecahan kata menjadi unit token yang lebih kecil apabila kata tersebut tidak ditemukan secara utuh pada *vocabulary* model. Pendekatan ini memungkinkan model menangani variasi morfologi bahasa Arab yang kompleks dan mengurangi masalah *out-of-vocabulary* token.

Selain melakukan tokenisasi, *tokenizer* juga menghasilkan representasi numerik berupa *input_ids*, *attention_mask*, dan *labels* yang diperlukan sebagai input model Transformer.

4.4.2 Tokenisasi dan *Label Alignment*

Pada tugas *token classification*, setiap token harus memiliki label kelas yang sesuai. Namun demikian, penggunaan *subword tokenization* menyebabkan satu kata dapat dipecah menjadi beberapa subword token. Oleh karena itu, dilakukan proses label alignment untuk menyesuaikan label kelas kata terhadap hasil tokenisasi AraBERT.

Pada penelitian ini, label asli hanya diberikan pada token pertama hasil tokenisasi suatu kata, sedangkan token *subword* berikutnya diberi label -100. Label -100 digunakan agar token tersebut diabaikan dalam perhitungan loss dan evaluasi model.

Sebagai contoh, suatu kata bahasa Arab dapat dipecah menjadi beberapa *subword* token oleh *tokenizer* AraBERT. Dalam kondisi tersebut, hanya token pertama yang digunakan sebagai target klasifikasi, sedangkan token lanjutan tidak dihitung sebagai target evaluasi. Pendekatan ini penting untuk menjaga kesesuaian antara jumlah token hasil tokenisasi dan jumlah label kelas kata pada dataset.

4.4.3 Representasi Input Model

Setelah proses tokenisasi dan label alignment selesai dilakukan, setiap sequence direpresentasikan ke dalam format input yang sesuai untuk model Transformer. Representasi tersebut terdiri atas *input_ids*, *attention_mask*, dan *labels*.

input_ids merupakan representasi numerik token berdasarkan *vocabulary* AraBERT. *attention_mask* digunakan untuk membedakan token valid dan *padding* token selama proses perhatian (*self-attention*) pada Transformer. Sementara itu,

labels digunakan sebagai target klasifikasi token pada proses pelatihan model. Kemudian seluruh *sequence* dipotong dan dipadding hingga panjang maksimum 128 token agar seluruh input memiliki ukuran yang seragam selama proses pelatihan.

4.4.4 Inisialisasi Model *Token Classification*

Model yang digunakan adalah *AutoModelForTokenClassification* berbasis *pretrained* AraBERT. Arsitektur ini merupakan pengembangan dari model BERT yang ditambahkan *classification head* pada bagian output untuk melakukan prediksi kelas pada setiap token.

Jumlah label output disesuaikan dengan jumlah kelas pada penelitian ini, yaitu *isim*, *fi'il*, dan *harf*. Dengan demikian, model menghasilkan probabilitas prediksi kelas untuk setiap token pada *sequence input*.

Pada tahap ini, bobot awal model berasal dari hasil *pretraining* AraBERT pada corpus bahasa Arab berskala besar sehingga model telah memiliki representasi linguistik bahasa Arab sebelum dilakukan *fine-tuning*.

4.4.5 Konfigurasi Pelatihan dan *Hyperparameter*

Proses pelatihan model dilakukan menggunakan pustaka Hugging Face Transformers dengan konfigurasi *hyperparameter* yang sama pada seluruh skenario eksperimen. Konfigurasi *hyperparameter* yang digunakan pada proses pelatihan model disajikan pada Tabel 4.4.

Tabel 4.4 Konfigurasi Pelatihan Model

Parameter	Nilai
<i>Learning Rate</i>	3e-5
<i>Batch Size Training</i>	8
<i>Batch Size Validation</i>	8

Parameter	Nilai
<i>Maksimum Epoch</i>	50
<i>Weight Decay</i>	0.01
<i>Optimizer</i>	AdamW
<i>Evaluation Strategy</i>	<i>Epoch</i>
<i>Save Strategy</i>	<i>Epoch</i>
<i>Early Stopping Patience</i>	3
Metric Evaluasi Utama	<i>Validation Loss</i>

Learning rate sebesar $3e-5$ digunakan untuk menjaga kestabilan pembaruan bobot selama *fine-tuning*. *Optimizer* AdamW digunakan karena mampu menangani optimisasi parameter Transformer secara lebih stabil dibandingkan metode optimisasi konvensional. Jumlah maksimum pelatihan ditetapkan sebesar 50 *epoch* untuk memberikan ruang yang cukup bagi model mencapai kondisi konvergen selama proses pelatihan. Namun, pelatihan tidak selalu dijalankan hingga seluruh *epoch* selesai karena penelitian ini menerapkan mekanisme *early stopping* berdasarkan performa *validation loss*. Nilai *early stopping patience* sebesar 3 dipilih untuk memberikan toleransi terhadap fluktuasi *validation loss* antar *epoch*, sehingga model masih memiliki kesempatan melakukan perbaikan performa sebelum proses pelatihan dihentikan secara otomatis. Evaluasi model dilakukan pada setiap *epoch* menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* terhadap data validasi.

4.4.6 Proses Pelatihan dan Pembaruan Bobot

Pada tahap pelatihan, model menerima input berupa token bahasa Arab yang telah direpresentasikan ke dalam bentuk numerik. Selanjutnya dilakukan proses *forward pass* untuk menghasilkan prediksi kelas pada setiap token.

Prediksi model kemudian dibandingkan dengan label sebenarnya menggunakan fungsi *Cross-Entropy Loss* untuk menghitung tingkat kesalahan model. Nilai *loss* yang diperoleh digunakan pada proses *backpropagation* untuk menghitung *gradien* parameter model.

Selanjutnya *optimizer* AdamW memperbarui parameter bobot (*weights*) model secara bertahap agar kesalahan prediksi semakin kecil pada iterasi berikutnya. Melalui proses ini, representasi bahasa yang sebelumnya dipelajari oleh AraBERT selama pretraining disesuaikan secara khusus terhadap tugas klasifikasi kelas kata bahasa Arab pada penelitian ini.

Proses *forward pass*, perhitungan *loss*, *backpropagation*, dan pembaruan bobot dilakukan secara iteratif pada setiap batch data selama satu *epoch* pelatihan. Setelah seluruh *batch* selesai diproses, model kemudian dievaluasi menggunakan data validasi sebelum melanjutkan ke *epoch* berikutnya.

Hasil pelatihan menunjukkan bahwa model mengalami penurunan *training loss* secara konsisten selama proses *fine-tuning* berlangsung. Hal ini menunjukkan bahwa model mampu mempelajari hubungan antara token dan label kelas secara efektif.

4.4.7 Early Stopping dan Pemilihan Checkpoint

Penelitian ini menerapkan mekanisme *early stopping* dengan nilai *patience* sebesar 3 epoch untuk mencegah terjadinya *overfitting* selama proses pelatihan model. Nilai *patience* tersebut digunakan untuk memberikan toleransi terhadap fluktuasi performa validasi yang dapat terjadi pada beberapa epoch, sehingga

pelatihan tidak dihentikan terlalu cepat ketika model masih berpotensi mengalami peningkatan performa.

Pada penelitian ini, jumlah maksimum pelatihan ditetapkan sebesar 50 epoch. Namun demikian, berdasarkan hasil eksperimen, proses pelatihan model berhenti pada sekitar epoch ke-4 hingga epoch ke-5 karena nilai *validation loss* tidak lagi menunjukkan perbaikan yang signifikan. Kondisi ini mengindikasikan bahwa model telah mencapai kondisi konvergen dan mampu mempelajari pola distribusi data secara optimal tanpa memerlukan pelatihan yang terlalu panjang.

Pada beberapa eksperimen, proses pelatihan tetap berlanjut hingga epoch ke-5 meskipun performa terbaik telah diperoleh pada epoch sebelumnya. Hal ini terjadi karena mekanisme *early stopping* masih memberikan kesempatan kepada model untuk melakukan perbaikan performa dalam beberapa epoch berikutnya sesuai nilai *patience* yang digunakan. Ketika tidak ditemukan penurunan *validation loss* setelah beberapa epoch berturut-turut, proses pelatihan kemudian dihentikan secara otomatis.

Temuan tersebut sejalan dengan penelitian (Rogers et al., 2020) yang menjelaskan bahwa proses *fine-tuning* model BERT umumnya mengalami penyesuaian parameter paling besar pada beberapa epoch awal, terutama pada layer bagian atas model. Oleh karena itu, jumlah epoch pelatihan pada model Transformer umumnya relatif sedikit dan ditentukan berdasarkan performa validasi selama proses pelatihan berlangsung.

Selain itu, penelitian ini menggunakan mekanisme pemilihan *checkpoint* terbaik secara otomatis berdasarkan nilai *validation loss* terendah. Dengan

pendekatan ini, model dengan performa validasi terbaik akan disimpan dan digunakan sebagai model final meskipun proses pelatihan masih berjalan pada epoch berikutnya. Strategi tersebut membantu menjaga kemampuan generalisasi model terhadap data baru sekaligus mengurangi risiko penggunaan model yang mengalami *overfitting*.

4.5 Hasil Pelatihan dan Evaluasi Model

Subbab ini membahas hasil proses *fine-tuning* model AraBERT pada tiga skenario rasio pembagian data, yaitu 70:15:15, 80:10:10, dan 90:5:5. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, serta *validation loss* pada setiap epoch untuk menganalisis performa dan kemampuan generalisasi model.

4.5.1 Hasil Pelatihan Split 70:15:15

Pada skenario rasio 70:15:15, sebanyak 70% data digunakan sebagai data latih, sedangkan masing-masing 15% digunakan sebagai data validasi dan data uji. Hasil pelatihan model pada skenario ini ditunjukkan pada Tabel 4.5.

Tabel 4.5 Hasil Pelatihan Model Rasio Split 70:15:15

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1-score
1	0.055219	0.042664	0.987711	0.988001	0.987145	0.987545
2	0.039988	0.037521	0.989492	0.988487	0.990149	0.989309
3	0.022396	0.052932	0.985752	0.987283	0.983845	0.985534
4	0.014958	0.056469	0.990027	0.989714	0.990039	0.989869
5	0.016701	0.050918	0.990472	0.989678	0.991024	0.990347

Berdasarkan hasil eksperimen, model menunjukkan performa yang sangat baik dengan nilai *F1-score* terbaik sebesar 0.990347 pada epoch ke-5. Nilai *training*

loss mengalami penurunan secara bertahap dari 0.055219 menjadi 0.016701, yang menunjukkan bahwa model mampu mempelajari pola data pelatihan dengan baik.

Nilai *validation loss* terendah diperoleh pada epoch ke-2 sebesar 0.037521, kemudian mengalami fluktuasi pada epoch berikutnya. Meskipun demikian, nilai *accuracy*, *precision*, *recall*, dan *F1-score* tetap berada pada rentang yang tinggi sehingga model masih menunjukkan kemampuan generalisasi yang baik.

Skenario ini memiliki ukuran data validasi dan data uji yang lebih besar dibandingkan skenario lainnya sehingga evaluasi performa model menjadi lebih stabil dan representatif terhadap kondisi data nyata. Namun, jumlah data latih yang lebih sedikit menyebabkan performa model sedikit lebih rendah dibandingkan rasio *split* lainnya.

4.5.2 Hasil Pelatihan Split 80:10:10

Pada skenario rasio 80:10:10, proporsi data latih ditingkatkan menjadi 80%, sedangkan masing-masing 10% digunakan sebagai data validasi dan data uji. Hasil pelatihan model ditunjukkan pada Tabel 4.6.

Tabel 4.6 Hasil Pelatihan Model Rasio Split 80:10:10

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1-score
1	0.047172	0.039712	0.988176	0.986252	0.989647	0.987922
2	0.027415	0.043286	0.990700	0.991444	0.989562	0.990488
3	0.015339	0.045204	0.989770	0.988455	0.990496	0.989466
4	0.012617	0.049640	0.990833	0.991076	0.990009	0.990538

Hasil eksperimen pada Tabel 4.6 menunjukkan bahwa model memperoleh performa yang lebih baik dibandingkan skenario 70:15:15. Nilai *F1-score* terbaik mencapai 0.990538 pada epoch ke-4 dengan nilai *accuracy* sebesar 0.990833.

Penambahan jumlah data latih memungkinkan model mempelajari pola linguistik bahasa Arab secara lebih optimal. Hal ini terlihat dari penurunan *training loss* yang cukup stabil hingga mencapai 0.012617 pada epoch ke-4.

Selain itu, ukuran data validasi dan data uji pada skenario ini masih cukup representatif untuk mengevaluasi kemampuan generalisasi model. Oleh karena itu, rasio 80:10:10 memberikan keseimbangan yang baik antara jumlah data pelatihan dan reliabilitas evaluasi.

4.5.3 Hasil Pelatihan Split 90:5:5

Pada skenario rasio 90:5:5, sebagian besar data digunakan sebagai data latih sehingga model memperoleh jumlah data pembelajaran paling besar. Hasil pelatihan model ditunjukkan pada Tabel 4.7.

Tabel 4.7 Hasil Pelatihan Model Rasio Split 90:5:5

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1-score
1	0.065947	0.028917	0.991671	0.990527	0.992705	0.991603
2	0.024626	0.023849	0.992972	0.992897	0.992896	0.992895
3	0.016926	0.042627	0.991411	0.990672	0.992025	0.991344
4	0.006166	0.041762	0.993753	0.994108	0.993249	0.993673
5	0.013895	0.040650	0.992192	0.991369	0.992881	0.992118

Berdasarkan hasil eksperimen, model memperoleh performa tertinggi dibandingkan seluruh skenario dengan nilai *F1-score* mencapai 0.993673 pada epoch ke-4. Nilai *training loss* juga mengalami penurunan sangat signifikan hingga mencapai 0.006166.

Jumlah data latih yang sangat besar membuat model mampu mempelajari distribusi data secara lebih mendalam sehingga performa metrik evaluasi meningkat. Namun demikian, ukuran data validasi dan data uji pada skenario ini

relatif kecil sehingga hasil evaluasi berpotensi kurang stabil dan kurang merepresentasikan kemampuan generalisasi model pada data baru.

4.5.4 Analisis Hasil *Fine-Tuning*

Berdasarkan seluruh hasil eksperimen, ketiga skenario rasio split data menghasilkan performa yang sangat tinggi dengan nilai *F1-score* di atas 0.98. Hal ini menunjukkan bahwa model AraBERT mampu mempelajari pola morfologi dan konteks bahasa Arab dengan sangat baik pada tugas *token classification* kelas kata.

Peningkatan proporsi data latih cenderung meningkatkan performa model. Hal ini terlihat pada rasio 90:5:5 yang menghasilkan nilai *F1-score* tertinggi karena model memperoleh jumlah data pembelajaran paling besar.

Meskipun demikian, ukuran data validasi dan data uji yang terlalu kecil dapat menyebabkan hasil evaluasi menjadi kurang representatif. Pada skenario 90:5:5, nilai performa yang sangat tinggi berpotensi dipengaruhi oleh ukuran data evaluasi yang lebih sedikit sehingga kemampuan generalisasi model belum dapat dibuktikan secara optimal.

Sebaliknya, rasio 70:15:15 memiliki ukuran data validasi dan data uji yang lebih besar sehingga evaluasi performa lebih stabil. Akan tetapi, jumlah data latih yang lebih sedikit menyebabkan performa model sedikit lebih rendah dibandingkan dua skenario lainnya.

Rasio 80:10:10 menunjukkan keseimbangan yang paling baik antara jumlah data pelatihan dan reliabilitas evaluasi. Skenario ini mampu menghasilkan performa tinggi dengan ukuran data evaluasi yang masih cukup representatif untuk mengukur kemampuan generalisasi model.

Selain itu, seluruh eksperimen menunjukkan bahwa proses pelatihan berhenti pada epoch awal meskipun maksimum epoch ditetapkan sebanyak 50 *epoch*. Hal ini menunjukkan bahwa model telah mencapai kondisi *konvergen* pada beberapa *epoch* pertama sehingga pelatihan tambahan tidak memberikan peningkatan performa yang signifikan.

4.5.5 Pemilihan Model Terbaik

Berdasarkan hasil seluruh eksperimen, penelitian ini memilih model dengan rasio *split data* 80:10:10 sebagai model terbaik yang diintegrasikan ke dalam sistem chatbot pembelajaran bahasa Arab.

Meskipun rasio 90:5:5 menghasilkan nilai *F1-score* tertinggi, ukuran data validasi dan data uji pada skenario tersebut relatif kecil sehingga evaluasi performa berpotensi kurang representatif terhadap kemampuan generalisasi model pada data baru.

Sementara itu, rasio 80:10:10 memberikan keseimbangan yang lebih baik antara performa model dan *reliabilitas* evaluasi. Model pada skenario ini memperoleh nilai *F1-score* sebesar 0.990538 dengan performa validasi yang stabil serta ukuran data evaluasi yang masih cukup *representatif*.

Dengan demikian, model hasil *fine-tuning* pada rasio 80:10:10 dipilih sebagai model akhir karena dinilai paling optimal untuk diimplementasikan pada sistem chatbot pembelajaran bahasa Arab.

4.6 Evaluasi Model pada Data Uji

Setelah model terbaik diperoleh berdasarkan hasil evaluasi pada data validasi, dilakukan evaluasi akhir menggunakan data uji (*testing set*) yang tidak digunakan selama proses pelatihan maupun validasi. Evaluasi ini bertujuan untuk mengukur kemampuan generalisasi model terhadap data baru yang belum pernah dilihat sebelumnya.

Evaluasi dilakukan menggunakan fungsi *evaluate()* dan *predict()* pada dataset uji. Pada proses evaluasi, token dengan label -100 tidak disertakan dalam perhitungan metrik karena token tersebut merepresentasikan *padding token* dan *subword* yang diabaikan selama proses *token classification*. Dengan demikian, hasil evaluasi hanya merepresentasikan performa klasifikasi pada token yang valid.

Model yang digunakan pada tahap evaluasi merupakan model terbaik hasil *fine-tuning* dengan rasio split data 80:10:10 yang dipilih berdasarkan keseimbangan antara performa model dan reliabilitas evaluasi.

4.6.1 Metrik Evaluasi

Hasil evaluasi pada data uji ditunjukkan pada Tabel 4.8 berikut ini.

Tabel 4.8 Hasil Evaluasi pada Data Uji

Metrik	Nilai
<i>Loss</i>	0.041832
<i>Accuracy</i>	0.988493
<i>Precision</i>	0.986599
<i>Recall</i>	0.990314
<i>F1-score</i>	0.988426

Berdasarkan pada Tabel 4.8, hasil evaluasi, model memperoleh nilai *accuracy* sebesar 0.988493 atau sekitar 98,85%, yang menunjukkan bahwa sebagian besar token pada data uji berhasil diklasifikasikan dengan benar.

Nilai *precision*, *recall*, dan *F1-score* yang seluruhnya berada pada kisaran 0.98 menunjukkan bahwa model memiliki performa yang sangat baik dan seimbang dalam mengidentifikasi kelas kata bahasa Arab. Nilai *loss* yang relatif kecil juga menunjukkan bahwa prediksi model memiliki tingkat kesalahan yang rendah terhadap label sebenarnya.

Perbedaan nilai *F1-score* antara data validasi dan data uji tergolong sangat kecil. Hal ini menunjukkan bahwa model tidak mengalami penurunan performa yang signifikan ketika diuji pada data baru, sehingga model memiliki kemampuan generalisasi yang baik. Rincian performa model pada masing-masing kelas ditampilkan pada Tabel 4.9.

Tabel 4.9 *Classification Report* per Kelas

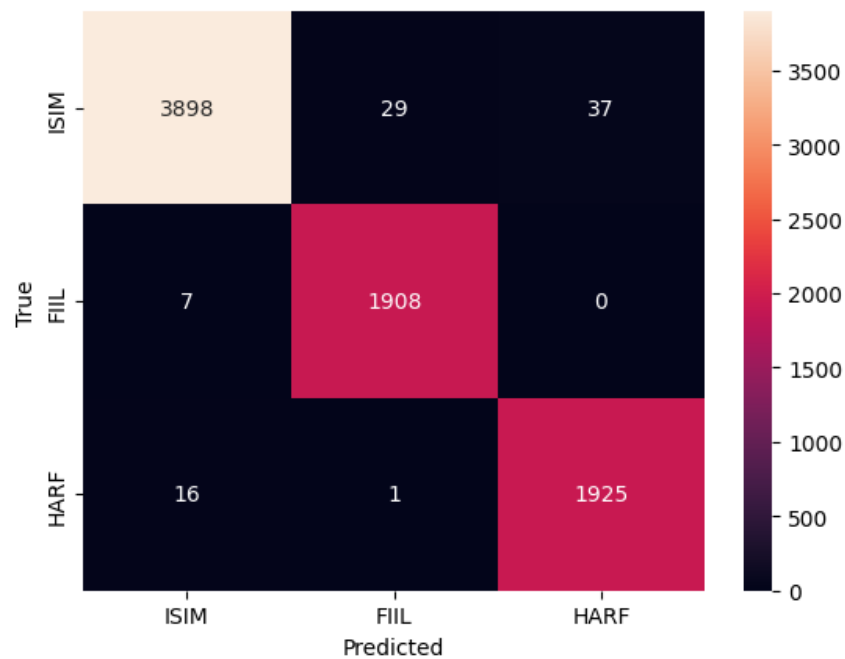
Kelas	Precision	Recall	F1-score	Support
ISIM	0.99	0.98	0.99	3964
FI'IL	0.98	1.00	0.99	1915
HARF	0.98	0.99	0.99	1942

Berdasarkan hasil *classification report*, seluruh kelas memperoleh nilai *precision*, *recall*, dan *F1-score* yang sangat tinggi dengan performa relatif seimbang antar kelas. Kelas *isim* memiliki jumlah data (*support*) terbesar, yaitu sebanyak 3964 token. Model mampu mengklasifikasikan kelas *isim* dengan nilai *F1-score* sebesar 0.99, yang menunjukkan bahwa model dapat mengenali karakteristik nominal bahasa Arab dengan sangat baik.

Pada kelas *fi'il*, model memperoleh nilai *recall* sebesar 1.00 yang menunjukkan bahwa hampir seluruh token *fi'il* berhasil dikenali dengan benar oleh model. Sementara itu, kelas HARF juga menunjukkan performa yang sangat baik dengan nilai *F1-score* sebesar 0.99.

4.6.2 Analisis Confusion Matrix

Selain menggunakan metrik evaluasi numerik, penelitian ini juga menganalisis pola kesalahan klasifikasi menggunakan *confusion matrix*. Visualisasi *confusion matrix* digunakan untuk melihat distribusi prediksi benar dan kesalahan prediksi pada masing-masing kelas. Hasil visualisasi confusion matrix ditunjukkan pada Gambar 4.1.



Gambar 4.1 Confusion Matrix pada Data Uji

Berdasarkan hasil *confusion matrix* pada Gambar 4.1, sebagian besar prediksi berada pada diagonal utama matriks. Hal ini menunjukkan bahwa mayoritas token berhasil diklasifikasikan dengan benar oleh model hasil *fine-tuning*.

Pada kelas *isim*, model berhasil mengklasifikasikan dengan benar sebanyak 3898 token dari total 3964 token. Kesalahan klasifikasi relatif kecil, yaitu sebanyak 29 token diprediksi sebagai *fi'il* dan 37 token diprediksi sebagai *harf*.

Pada kelas *fi'il*, model menghasilkan 1908 prediksi benar dari total 1915 token. Kesalahan yang terjadi sangat sedikit, yaitu 7 token salah diprediksi sebagai *isim* dan tidak ada token yang terprediksi sebagai *harf*. Hasil ini menunjukkan bahwa model mampu mengenali pola verba bahasa Arab dengan sangat baik.

Sementara itu, pada kelas *harf*, model berhasil mengklasifikasikan dengan benar sebanyak 1925 token dari total 1942 token. Terdapat 16 token yang salah diprediksi sebagai *isim* dan 1 token sebagai *fi'il*.

Secara umum, jumlah kesalahan klasifikasi pada seluruh kelas tergolong sangat kecil dibandingkan total token yang diuji. Pola kesalahan yang paling sering muncul terjadi antara kelas *isim* dan *harf*. Hal ini dapat dipengaruhi oleh adanya kemiripan fungsi sintaktis maupun karakteristik morfologis pada beberapa konteks bahasa Arab.

Dominasi nilai pada diagonal utama *confusion matrix* menunjukkan bahwa model hasil *fine-tuning* mampu membentuk representasi kontekstual yang baik dalam membedakan kelas *isim*, *fi'il*, dan *harf* pada data uji.

Hasil evaluasi ini juga konsisten dengan nilai *accuracy* dan *F1-score* yang mendekati 0,99, sehingga menunjukkan bahwa model memiliki kemampuan generalisasi yang stabil terhadap data baru yang belum pernah dilihat sebelumnya.

4.6.3 Perbandingan dengan Baseline Model

Untuk mengetahui pengaruh proses *fine-tuning* terhadap performa klasifikasi, penelitian ini membandingkan model hasil *fine-tuning* dengan *baseline model* berupa AraBERT pralatih (*pretrained model*) tanpa proses pelatihan ulang pada dataset penelitian. Pada skenario ini, model digunakan secara langsung untuk

melakukan prediksi pada data uji tanpa pembaruan parameter terhadap tugas *token classification* kelas kata bahasa Arab. Hasil evaluasi *baseline model* ditunjukkan pada Tabel 4.10 berikut.

Tabel 4.10 Hasil Evaluasi *Baseline Model* (Tanpa *Fine-Tuning*)

Metrik	Nilai
<i>Loss</i>	1,2670
<i>Accuracy</i>	0,2714
<i>Precision</i>	0,3407
<i>Recall</i>	0,3121
<i>F1-score</i>	0,2686

Berdasarkan hasil evaluasi pada Tabel 4.10, terlihat bahwa performa *baseline model* masih sangat rendah dengan nilai *F1-score* sebesar 0,2686 dan *accuracy* sebesar 27,14%. Nilai *loss* yang mencapai 1,2670 juga menunjukkan bahwa prediksi model masih jauh dari label sebenarnya.

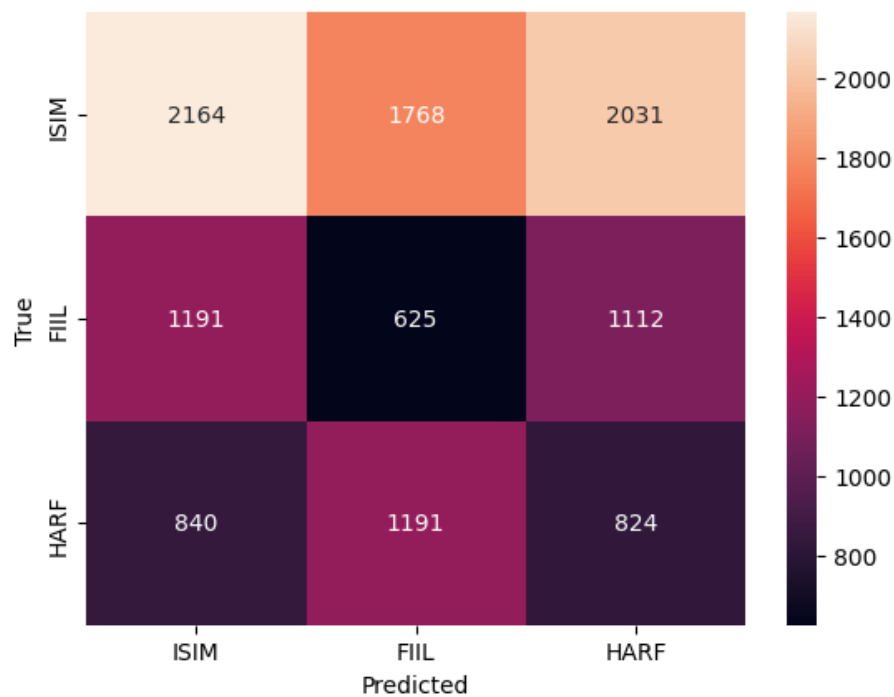
Rendahnya performa tersebut terjadi karena model AraBERT pralatih belum pernah disesuaikan dengan tugas klasifikasi kelas kata bahasa Arab pada penelitian ini. Pada saat model dimuat tanpa proses *fine-tuning*, parameter pada *classification layer* masih berada pada kondisi inisialisasi awal (*newly initialized*) dan belum mengalami pembelajaran terhadap label *isim*, *fi'il*, dan *harf*.

Hal ini ditunjukkan pada proses pemuatan model yang menghasilkan status *classifier.weight* dan *classifier.bias* sebagai *missing parameters*, yang berarti parameter klasifikasi tidak tersedia pada model pralatih dan harus dipelajari melalui proses *fine-tuning*. Akibatnya, model belum mampu membentuk pemetaan yang baik antara representasi token dan kelas kata yang digunakan dalam penelitian.

Jika dibandingkan dengan model hasil *fine-tuning* yang memperoleh *F1-score* sebesar 0,9884, terlihat adanya peningkatan performa yang sangat signifikan.

Temuan ini menunjukkan bahwa proses *fine-tuning* berperan penting dalam menyesuaikan representasi kontekstual AraBERT terhadap karakteristik linguistik pada tugas *token classification* bahasa Arab.

Untuk menganalisis pola kesalahan klasifikasi pada *baseline model*, digunakan *confusion matrix* yang ditampilkan pada Gambar 4.2.



Gambar 4.2 Confusion Matrix *Baseline Model* (Tanpa *Fine-Tuning*)

Berdasarkan confusion matrix pada Gambar 4.2, terlihat bahwa distribusi prediksi model masih sangat tidak stabil dan banyak token diklasifikasikan ke kelas yang salah. Pada kelas *isim*, model hanya mampu memprediksi benar sebanyak 2164 token dari total 5963 token. Sisanya salah diklasifikasikan sebagai *fi'il* sebanyak 1768 token dan *harf* sebanyak 2031 token. Hal ini menunjukkan bahwa model belum mampu mengenali karakteristik linguistik kelas *isim* secara konsisten.

Pada kelas *fi'il*, model hanya menghasilkan 625 prediksi benar dari total 2928 token. Sebagian besar token lainnya salah diprediksi sebagai *isim* sebanyak 1191 token dan *harf* sebanyak 1112 token. Kondisi ini menunjukkan bahwa model mengalami kesulitan dalam membedakan pola morfologis verba bahasa Arab tanpa proses adaptasi melalui *fine-tuning*.

Sementara itu, pada kelas *harf*, model hanya mampu memprediksi benar sebanyak 824 token dari total 2855 token. Kesalahan klasifikasi terjadi cukup besar ke kelas *isim* sebanyak 840 token dan ke kelas *fi'il* sebanyak 1191 token. Hal ini menunjukkan bahwa model belum mampu memahami fungsi sintaktis partikel secara memadai pada tugas klasifikasi token.

Secara keseluruhan, nilai pada diagonal utama *confusion matrix* tidak dominan, yang menandakan bahwa *baseline model* gagal membentuk pola klasifikasi yang konsisten. Kondisi tersebut berbeda jauh dibandingkan model hasil *fine-tuning* yang menunjukkan dominasi prediksi benar pada diagonal utama matriks.

Hasil perbandingan ini memperlihatkan bahwa proses *fine-tuning* memberikan peningkatan performa yang sangat signifikan terhadap kemampuan model dalam melakukan klasifikasi kelas kata bahasa Arab. Dengan demikian, proses *fine-tuning* terbukti efektif dalam menyesuaikan representasi bahasa AraBERT terhadap kebutuhan tugas *token classification* pada penelitian ini.

4.7 Implementasi Chatbot Pembelajaran Bahasa Arab

Implementasi model AraBERT hasil *fine-tuning* pada penelitian ini berupa pengembangan sebuah chatbot pembelajaran bahasa Arab yang mampu melakukan

analisis kelas kata (*Part-of-Speech*) ke dalam kategori *isim* (kata benda), *fi'il* (kata kerja), dan *harf* (partikel) terhadap kalimat masukan pengguna.

Untuk mendukung penggunaan oleh pengguna pemula, sistem chatbot dirancang untuk dapat menerima input dalam bahasa Indonesia. Hal ini bertujuan untuk memudahkan pengguna yang belum terbiasa mengetik atau menyusun kalimat dalam bahasa Arab. Oleh karena itu, sistem dilengkapi dengan tahap tambahan berupa proses penerjemahan (*translation*) dari bahasa Indonesia ke bahasa Arab sebelum dilakukan proses klasifikasi kelas kata oleh model AraBERT.

Chatbot ini dirancang sebagai sistem inferensi berbasis *Natural Language Processing* (NLP) yang memanfaatkan model token classification hasil pelatihan sebagai inti pengambilan keputusan. Seluruh keluaran sistem klasifikasi merupakan hasil prediksi langsung dari model Transformer yang telah dilatih menggunakan dataset teks Qur'ani, sedangkan proses penerjemahan berfungsi sebagai tahap pendukung untuk menjembatani input pengguna dengan model. Implementasi sistem dilakukan dalam bentuk aplikasi berbasis web lokal menggunakan framework Streamlit. Proses inferensi model dapat dijalankan secara lokal, namun fitur penerjemahan memerlukan koneksi internet karena menggunakan layanan penerjemahan eksternal.

4.7.1 Tahapan Proses Sistem Chatbot Pembelajaran

Sistem chatbot yang diimplementasikan pada penelitian ini terdiri atas beberapa tahapan proses yang menghubungkan input pengguna dengan keluaran analisis kelas kata secara otomatis. Sistem dirancang untuk menerima input dalam bentuk teks bahasa Arab maupun bahasa Indonesia.

Proses diawali dengan penerimaan input pengguna dan deteksi bahasa secara otomatis. Jika input berupa bahasa Indonesia, sistem akan menerjemahkan teks ke dalam bahasa Arab sebelum diproses lebih lanjut. Setelah itu, teks melalui tahap normalisasi dan tokenisasi sebelum dianalisis menggunakan model AraBERT hasil *fine-tuning*.

Model kemudian melakukan klasifikasi kelas kata pada setiap token ke dalam kategori *isim*, *fi'il*, dan *harf*. Hasil prediksi selanjutnya diproses kembali melalui tahap post-processing untuk menggabungkan subword token menjadi kata utuh agar lebih mudah dipahami pengguna.

Pada tahap akhir, sistem menampilkan hasil klasifikasi kelas kata beserta informasi tambahan berupa arti kata dalam bahasa Indonesia sebagai bagian dari implementasi chatbot pembelajaran bahasa Arab.

4.7.2 Proses Inferensi dan Post-processing

Proses inferensi dalam sistem chatbot dilakukan menggunakan model AraBERT fine-tuned dan tokenizer yang telah disimpan secara lokal pada direktori proyek. Model dimuat menggunakan parameter *local_files_only=True* sehingga proses utama inferensi dapat dijalankan secara lokal tanpa memerlukan pelatihan ulang. Namun, koneksi internet tetap diperlukan pada tahap penerjemahan bahasa Indonesia ke bahasa Arab karena sistem memanfaatkan layanan GoogleTranslator sebagai modul translasi eksternal.

Pada tahap awal, sistem menerima input teks dari pengguna dan melakukan deteksi bahasa secara otomatis. Jika input teridentifikasi sebagai bahasa Indonesia, teks terlebih dahulu diterjemahkan ke bahasa Arab menggunakan

GoogleTranslator. Hasil terjemahan tersebut kemudian diproses melalui tahap normalisasi untuk menjaga kesesuaian format dengan data pelatihan model.

Selanjutnya, tokenizer AraBERT mengubah teks menjadi token-token subword dan merepresentasikannya dalam bentuk *input_ids* serta *attention_mask*. Representasi ini menjadi masukan utama bagi model AraBERT *fine-tuned* untuk menghasilkan prediksi label kelas kata pada setiap token.

Keluaran model pada level token kemudian diproses melalui tahap *post-processing* dengan memanfaatkan pemetaan token terhadap kata asli (*word alignment*) untuk menggabungkan subword token menjadi kata utuh. Tahap ini penting agar hasil klasifikasi lebih interpretatif dalam konteks pembelajaran.

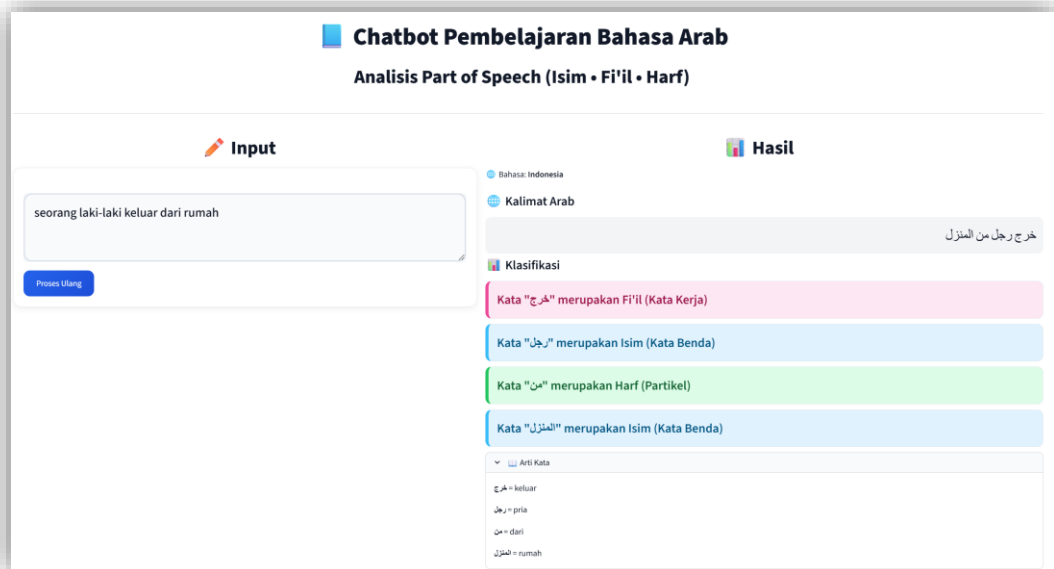
Setelah proses penggabungan selesai, setiap kata ditampilkan bersama label kelas katanya dalam bentuk edukatif, yaitu *isim* (kata benda), *fi'il* (kata kerja), dan *harf* (partikel). Sistem juga dapat menambahkan informasi arti kata dalam bahasa Indonesia sebagai dukungan pembelajaran tambahan. Dengan demikian, proses inferensi tidak hanya menghasilkan prediksi linguistik, tetapi juga menyajikan keluaran yang lebih mudah dipahami oleh pengguna sebagai bagian dari implementasi chatbot pembelajaran bahasa Arab.

4.7.3 Antarmuka dan Skenario Penggunaan

Antarmuka chatbot dikembangkan menggunakan framework Streamlit berbasis web lokal dan dijalankan melalui perintah *streamlit run app.py*. Sistem dirancang sederhana dan interaktif dengan dua kondisi tampilan utama, yaitu halaman input terpusat sebelum analisis serta tampilan dua kolom setelah proses analisis dilakukan.

Pengguna dapat memasukkan kalimat dalam bahasa Indonesia maupun bahasa Arab melalui kolom input yang tersedia. Sistem secara otomatis mendeteksi bahasa input. Jika input berupa bahasa Indonesia, sistem akan menerjemahkannya terlebih dahulu ke dalam bahasa Arab sebelum dilakukan analisis, sedangkan input bahasa Arab diproses secara langsung.

Hasil analisis yang ditampilkan meliputi klasifikasi kelas kata (isim, fi'il, dan harf), hasil terjemahan bahasa Arab, serta arti kata dalam bahasa Indonesia yang disajikan secara interaktif. Setiap kelas kata divisualisasikan menggunakan warna berbeda untuk meningkatkan keterbacaan dan mempermudah identifikasi visual. Contoh tampilan antarmuka sistem ditunjukkan pada Gambar 4.3 dan Gambar 4.4.



Gambar 4.3 Tampilan Antarmuka Chatbot dengan Input Bahasa Indonesia

Gambar 4.3 menunjukkan tampilan sistem ketika pengguna memasukkan kalimat dalam bahasa Indonesia. Pada kondisi ini, sistem terlebih dahulu

melakukan proses penerjemahan ke dalam bahasa Arab sebelum menampilkan hasil klasifikasi kelas kata untuk setiap token.



Gambar 4.4 Tampilan Antarmuka Chatbot dengan Input Bahasa Arab

Gambar 4.4 menunjukkan tampilan sistem ketika pengguna memasukkan kalimat dalam bahasa Arab. Pada kondisi ini, sistem langsung melakukan proses analisis tanpa melalui tahap penerjemahan.

Dalam proses pengujian, dilakukan simulasi dengan berbagai input kalimat dalam bahasa Indonesia dan bahasa Arab. Sebagai contoh, pengguna dapat memasukkan kalimat dalam bahasa Indonesia seperti “*seorang laki-laki keluar dari rumah*”, yang kemudian secara otomatis diterjemahkan ke dalam bahasa Arab sebelum dilakukan proses klasifikasi.

Selain itu, sistem juga diuji menggunakan input langsung dalam bahasa Arab, seperti “الطالب يقرأ الكتاب”, di mana sistem akan langsung melakukan analisis tanpa melalui proses terjemahan.

Hasil analisis pada kedua skenario tersebut ditampilkan dalam bentuk klasifikasi kelas kata untuk setiap token serta arti kata dalam bahasa Indonesia. Penyajian hasil ini memberikan umpan balik visual yang membantu pengguna dalam memahami struktur kelas kata dalam bahasa Arab.

Hasil klasifikasi yang ditampilkan chatbot tidak hanya berfungsi sebagai output prediksi model, tetapi juga mendukung proses pembelajaran dasar bahasa Arab bagi pengguna. Melalui hasil klasifikasi token berupa *isim*, *fi'il*, dan *harf*, pengguna dapat memahami fungsi dasar setiap kata dalam suatu kalimat bahasa Arab secara lebih kontekstual.

Sebagai contoh, pengguna dapat mengenali bahwa kata yang menunjukkan pelaku, benda, atau konsep umumnya diklasifikasikan sebagai *isim*, sedangkan kata yang menunjukkan aktivitas atau tindakan diklasifikasikan sebagai *fi'il*. Adapun kata yang berfungsi sebagai penghubung atau partikel diklasifikasikan sebagai *harf*. Dengan demikian, hasil analisis yang diberikan chatbot tidak hanya menampilkan klasifikasi token, tetapi juga membantu pengguna mengenali pola dasar struktur gramatikal bahasa Arab.

Pendekatan ini mendukung proses pembelajaran morfologi dasar bahasa Arab, khususnya bagi pengguna pemula yang masih mempelajari pengenalan jenis kata dalam kalimat. Umpan balik otomatis yang diberikan sistem memungkinkan pengguna memperoleh proses belajar yang lebih interaktif dan praktis dibandingkan identifikasi kelas kata secara manual.

4.7.4 Evaluasi Hasil Klasifikasi Chatbot

Evaluasi implementasi dilakukan dengan menguji chatbot menggunakan sejumlah kalimat bahasa Arab baku (*Modern Standard Arabic/MSA*) dalam bentuk kalimat utuh. Pengujian ini bertujuan untuk memastikan bahwa model yang telah dilatih tetap bekerja secara konsisten setelah diintegrasikan ke dalam sistem chatbot, serta mengamati performanya dalam konteks penggunaan nyata.

Evaluasi utama dilakukan menggunakan kalimat bahasa Arab karena model yang digunakan merupakan model klasifikasi berbasis bahasa Arab, sehingga pengujian perlu dilakukan secara langsung tanpa melalui proses penerjemahan agar hasil tidak terpengaruh oleh kesalahan dari komponen eksternal. Sementara itu, penggunaan input bahasa Indonesia digunakan untuk merepresentasikan skenario penggunaan sistem secara nyata, di mana pengguna tidak harus memahami bahasa Arab. Oleh karena itu, pengujian dengan bahasa Indonesia hanya digunakan sebagai evaluasi tambahan pada level sistem (*end-to-end*), bukan sebagai evaluasi utama model.

Meskipun pengujian chatbot juga menggunakan kalimat bahasa Arab, pengujian ini memiliki tujuan yang berbeda dengan evaluasi model pada data uji di Subbab 4.5. Evaluasi model pada Subbab 4.5 dilakukan secara kuantitatif menggunakan dataset berukuran besar beserta metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Adapun evaluasi chatbot pada subbab ini bersifat kualitatif dan terbatas pada sejumlah contoh kalimat untuk mengamati konsistensi performa model setelah diintegrasikan ke dalam sistem.

Label aktual (ground truth) pada evaluasi ini disusun secara manual berdasarkan kaidah nahwu dasar yang membagi kalam ke dalam tiga kategori utama, yaitu isim, fi'il, dan harf. Penentuan label mengacu pada kaidah klasifikasi kata dalam literatur nahwu klasik, khususnya pembahasan mengenai aqşam al-kalam (pembagian kalam), tanda isim, tanda fi'il, dan tanda huruf sebagaimana dijelaskan dalam *Al-Tuhfah al-Saniyah Syarah al-Muqaddimah al-Ajrumiyyah* dan *Alfiyyah Ibnu Malik* (Abdul Hamid, 2019; Malik, 2006). Pemilihan referensi tersebut didasarkan pada kedudukannya sebagai rujukan dasar dalam kajian nahwu mengenai klasifikasi jenis kata bahasa Arab. Penentuan kelas kata dalam bahasa Arab juga memerlukan acuan tata bahasa Arab karena kompleksitas morfologi dan sintaksis bahasa Arab (Zeroual et al., 2017). Oleh karena itu, penyusunan label aktual pada penelitian ini dilakukan secara manual berdasarkan kaidah nahwu klasik untuk digunakan sebagai acuan evaluasi keluaran chatbot.

Dalam proses anotasi, penentuan label aktual tidak dilakukan hanya berdasarkan tanda singkat atau satu ciri gramatikal tertentu, melainkan melalui analisis nahwu secara bertahap terhadap setiap token. Tahap pertama adalah identifikasi makna dasar kata, yaitu dengan menentukan apakah suatu token menunjukkan benda, konsep, atau entitas yang berdiri sendiri (*isim*); menunjukkan pekerjaan, peristiwa, atau aktivitas yang berkaitan dengan waktu tertentu (*fi'il*); atau merupakan kata tugas yang maknanya baru sempurna ketika digabungkan dengan kata lain (*harf*).

Tahap kedua adalah identifikasi ciri morfologis dan gramatikal (*alamat nahwu*) secara lebih komprehensif. Pada kategori isim, analisis mencakup

kemungkinan token menerima alif-lam (ال), tanwin, huruf *jar*, *i'rab jar* (kasrah), serta bentuk-bentuk nominal lain seperti dhamir, *isim isyarah*, atau *isim maushul* jika relevan. Pada kategori *fi'il*, analisis mempertimbangkan pola *fi'il madhi*, *fi'il mudhari'*, *fi'il amr*, keberadaan awalan *mudhari'* (ي، ت، أ، ن), serta indikator seperti *qad* (قد), *sin* (س), *saufa* (سوف), dan *ta' ta'nits*. Adapun kategori harf diidentifikasi melalui ketiadaan ciri-ciri *isim* maupun *fi'il* serta fungsinya sebagai kata penghubung atau penanda relasi gramatikal.

Tahap ketiga adalah analisis konteks sintaktis token dalam susunan kalimat, sehingga klasifikasi tidak hanya bergantung pada bentuk kata secara terpisah, tetapi juga pada fungsi gramatikalnya dalam struktur kalimat. Sebagai contoh, pada kalimat الطالب يقرأ الكتاب, token الطالب diklasifikasikan sebagai isim karena menunjukkan subjek nominal dan menerima alif-lam, token يقرأ diklasifikasikan sebagai *fi'il* karena berpola *fi'il mudhari'* dan menunjukkan aktivitas membaca, sedangkan token الكتاب diklasifikasikan sebagai *isim* karena berfungsi sebagai objek nominal.

Dengan demikian, proses pelabelan aktual dilakukan berdasarkan kombinasi antara makna *leksikal*, ciri morfologis, serta fungsi sintaktis, sehingga anotasi yang dihasilkan tidak semata-mata bergantung pada tanda dasar, tetapi lebih sesuai dengan penerapan kaidah nahwu secara komprehensif. Ringkasan dasar kaidah yang digunakan dalam proses ini disajikan pada Tabel 4.11.

Tabel 4.11 Dasar Kaidah Penentuan Label Aktual Berdasarkan Nahwu

Kategori	Definisi Dasar	Tanda Utama	Penjelasan
Isim	Kata yang menunjukkan makna	Dapat menerima alif-lam (ال), tanwin,	Contoh: الطالب، كتاب، المدرسة

Kategori	Definisi Dasar	Tanda Utama	Penjelasan
	mandiri dan tidak terikat waktu	huruf jar, atau dibaca jar (kasrah)	
Fi'il	Kata yang menunjukkan pekerjaan/peristiwa yang terikat waktu	Dapat dimasuki qad (قد), sin (س), saufa (سوف), ta' ta'nits, atau menunjukkan makna perintah	Mencakup <i>fi'il madhi, mudhari'</i> , dan <i>amr</i>
Harf	Kata yang maknanya muncul jika digabung dengan kata lain	Tidak menerima tanda isim maupun fi'il	Contoh: في، من، إلى

Berdasarkan kaidah pada Tabel 4.11, proses anotasi dilakukan secara manual terhadap setiap token pada kalimat uji, kemudian dibandingkan dengan hasil prediksi chatbot. Hasil prediksi token secara rinci disajikan pada Tabel 4.12. Tabel ini menampilkan setiap token beserta label aktual dan label prediksi yang dihasilkan oleh sistem, sehingga transparansi keluaran model dapat diamati secara langsung.

Tabel 4.12 Hasil Pengujian Implementasi Chatbot pada Input Bahasa Arab

No	Kalimat	Kata	Label Aktual	Label Prediksi
1.	الطالب يقرأ الكتاب ("Siswa itu membaca buku")	الطالب	Isim	Isim
		يقرأ	Fi'il	Fi'il
		الكتاب	Isim	Isim
2.	المُعلِّم في الفصل ("Guru berada di kelas")	المعلم	Isim	Isim
		في	Harf	Harf
		الفصل	Isim	Isim
3.	خرج الرجل من البيت ("Laki-laki itu keluar dari rumah")	خرج	Fi'il	Fi'il
		الرجل	Isim	Isim
		من	Harf	Harf
		البيت	Isim	Isim
4.	الطفل الصغير يلعب مع أخيه في البيت ("Anak kecil itu bermain bersama saudaranya di rumah")	الطفل	Isim	Isim
		الصغير	Isim	Isim
		يلعب	Fi'il	Fi'il
		مع	Harf	Isim

		أخيه	Isim	Isim
		في	Harf	Harf
		البيت	Isim	Isim
5.	السَّمَاءُ صَافِيَةٌ الْيَوْمَ ("Langit cerah hari ini")	السَّمَاءُ	Isim	Isim
		صَافِيَةٌ	Isim	Isim
		اليوم	Isim	Isim
6.	اشْتَرَيْتُ قَلَمًا جَمِيلًا مِنَ السُّوقِ ("Saya membeli pena yang indah dari pasar")	اشْتَرَيْتُ	Fi'il	Fi'il
		قَلَمًا	Isim	Isim
		جَمِيلًا	Isim	Isim
		مِنَ	Harf	Harf
		السُّوقِ	Isim	Isim
7.	الْمُعَلِّمُ يَشْرَحُ الدَّرْسَ ("Guru menjelaskan pelajaran")	الْمُعَلِّمُ	Isim	Isim
		يَشْرَحُ	Fi'il	Fi'il
		الدَّرْسَ	Isim	Isim
8.	جَلَسَتِ الْبِنْتُ عَلَى الْكُرْسِيِّ ("Anak perempuan itu duduk di atas kursi")	جَلَسَ	Fi'il	Fi'il
		الْبِنْتُ	Isim	Isim
		عَلَى	Harf	Harf
		الْكُرْسِيِّ	Isim	Isim
9.	قَرَأَ مُحَمَّدٌ الصَّحِيفَةَ ("Muhammad membaca surat kabar")	قَرَأَ	Fi'il	Fi'il
		مُحَمَّدٌ	Isim	Isim
		الصَّحِيفَةَ	Isim	Isim
10.	الطَّالِبُ الْمُجْتَهِدُ يَقْرَأُ الْكِتَابَ الْيَوْمَ ("Siswa yang rajin itu membaca buku hari ini")	الطَّالِبُ	Isim	Isim
		الْمُجْتَهِدُ	Isim	Isim
		يَقْرَأُ	Fi'il	Fi'il
		الْكِتَابَ	Isim	Isim
		الْيَوْمَ	Isim	Isim
11.	ذَهَبَ الطَّالِبُ إِلَى الْمَكْتَبَةِ وَقَرَأَ الْكِتَابَ فِيهَا ("Siswa itu pergi ke perpustakaan dan membaca buku di sana")	ذَهَبَ	Fi'il	Fi'il
		الطَّالِبُ	Isim	Isim

		إلى	Harf	Harf
		المكتبة	Isim	Isim
		و	Harf	Fi'il
		قرأ	Fi'il	Fi'il
		الكتاب	Isim	Isim
		فيها	Harf	Harf
12.	سافرتُ إلى مصرَ للدراسة ثم عدتُ إلى بلدي ("Saya pergi ke Mesir untuk belajar kemudian kembali ke negaraku")	سافرت	Fi'il	Fi'il
		إلى	Harf	Harf
		مصر	Isim	Isim
		ل	Harf	Isim
		الدراسة	Isim	Isim
		ثم	Harf	Harf
		عدت	Fi'il	Fi'il
		إلى	Harf	Harf
		بلدي	Isim	Isim
13.	لم ينجح الطالبُ لأنَّهُ لم يذاكرَ جيّدًا ("Siswa itu tidak berhasil karena dia tidak belajar dengan baik")	لم	Harf	Harf
		ينجح	Fi'il	Fi'il
		الطالب	Isim	Isim
		لأن	Harf	Harf
		هـ	Isim	Harf
		لم	Harf	Harf
		يذاكر	Fi'il	Fi'il
		جيّدًا	Isim	Isim
14.	جلسَ ثم أكلَ تحتَ الشجرة ("Dia duduk kemudian makan di bawah pohon")	جلس	Fi'il	Fi'il
		ثم	Harf	Harf

Berdasarkan Tabel 4.12, mayoritas token berhasil diklasifikasikan sesuai dengan kategori gramatikalnya. Model menunjukkan konsistensi yang cukup baik

dalam mengidentifikasi *Fi'il* serta *Isim*, baik pada struktur kalimat sederhana maupun kombinasi antar kelas kata dalam satu kalimat.

Kesalahan klasifikasi relatif terbatas dan umumnya melibatkan token berbentuk partikel pendek atau klitik, termasuk kategori *Harf* serta pronomina terikat (*dhamir muttashil*). Pola kesalahan terutama ditemukan pada token seperti *و* dan *ل* yang melekat pada kata lain, bentuk gabungan seperti *لِلدِّراسة*, serta huruf jar dalam konteks nominal seperti *مع*. Selain itu, kesalahan juga muncul pada pronomina terikat seperti “هـ” dalam bentuk *لأنه*. Untuk menganalisis pola tersebut secara lebih sistematis, jenis kesalahan yang muncul dirangkum pada Tabel 4.13.

Tabel 4.13 Klasifikasi Jenis Kesalahan

No	Jenis Kesalahan	Contoh Token	Label Aktual	Label Prediksi	Pola Kesalahan	Kemungkinan Penyebab
1.	Harf → Isim	مع	Harf	Isim	Huruf jar diapit dua isim	Model lebih menangkap pola konteks nominal
2.	Harf → Fi'il	و	Harf	Fi'il	Prefiks koordinatif sebelum fi'il	Token pendek tanpa segmentasi klitik
3.	Harf → Isim	ل (pada <i>لِلدِّراسة</i>)	Harf	Isim	Prefiks jar melekat pada isim	Tidak dilakukan segmentasi morfologis klitik
4.	Isim → Harf	هـ (dari <i>لأنه</i>)	Isim	Harf	Dhamir melekat setelah partikel	Model kesulitan pada token sangat pendek

Berdasarkan Tabel 4.13, kesalahan klasifikasi yang terjadi cenderung terkonsentrasi pada token berukuran pendek dan berbentuk klitik, seperti partikel *و* dan *ل* yang melekat pada kata lain. Token dengan panjang karakter minimal memiliki representasi yang sangat bergantung pada konteks, sehingga meningkatkan potensi kesalahan prediksi.

Selain itu, kasus seperti *للدراصة* menunjukkan bahwa model memproses token sebagai satu kesatuan tanpa memisahkan prefiks *jar* dari kata dasarnya. Hal ini menyebabkan label cenderung mengikuti kelas kata dominan, yaitu Isim, dibandingkan komponen prefiksnya yang termasuk Harf. Pola serupa juga ditemukan pada pronomina terikat seperti “هـ” dalam bentuk *لأنه* yang memiliki ambiguitas tinggi.

Secara umum, kesalahan ini berkaitan dengan keterbatasan pendekatan token-level classification yang tidak melakukan segmentasi morfologis secara eksplisit. Meskipun demikian, jumlah kesalahan relatif kecil dibandingkan total token yang diuji, sehingga model tetap menunjukkan performa yang stabil dalam mengklasifikasikan struktur kelas kata bahasa Arab.

Berdasarkan hasil pengujian, total token yang dianalisis berjumlah 67 token yang tersebar dalam 15 kalimat, dengan 62 token berhasil diklasifikasikan sesuai label aktual dan 5 token mengalami kesalahan prediksi. Tingkat akurasi klasifikasi token pada evaluasi implementasi chatbot dihitung menggunakan persamaan berikut:

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Total\ Token} = 100\%$$

$$Akurasi = \frac{62}{67} \times 100\% = 92.54\%$$

Hasil tersebut menunjukkan bahwa model *fine-tuning* AraBERT yang telah diintegrasikan ke dalam sistem chatbot mampu mempertahankan performa klasifikasi yang cukup stabil pada konteks kalimat utuh, tidak hanya pada data uji terpisah. Selain itu, hasil ini juga menunjukkan bahwa model tetap mampu

menghasilkan prediksi yang konsisten setelah diintegrasikan ke dalam sistem chatbot pada konteks penggunaan nyata.

Meskipun dataset pelatihan berasal dari anotasi morfologi ayat-ayat Al-Qur'an, pengujian implementasi pada penelitian ini dilakukan menggunakan kalimat bahasa Arab baku (MSA) untuk menguji kemampuan generalisasi model pada domain yang lebih luas. Pengujian khusus pada teks Al-Qur'an secara langsung dapat menjadi pengembangan lanjutan pada penelitian berikutnya.

Selain pengujian menggunakan kalimat bahasa Arab, dilakukan pula pengujian terbatas menggunakan input bahasa Indonesia untuk melihat kinerja sistem secara menyeluruh (*end-to-end*), termasuk tahap penerjemahan sebelum proses klasifikasi. Pengujian ini bertujuan untuk memastikan bahwa sistem tetap dapat bekerja dengan baik ketika digunakan sesuai dengan skenario pengguna, yaitu memasukkan kalimat dalam bahasa Indonesia.

Pengujian pada tahap implementasi ini dilakukan secara terbatas dan tidak dimaksudkan untuk mengukur performa model secara statistik, melainkan untuk mengamati konsistensi perilaku model pada konteks kalimat utuh setelah integrasi ke dalam sistem. Hasil pengujian tersebut disajikan pada Tabel 4.14.

Tabel 4.14 Hasil Pengujian Implementasi Chatbot pada Input Bahasa Indonesia

No	Kalimat Indonesia	Hasil Terjemahan (Arab)	Token	Label Aktual	Label Prediksi
1.	Seorang laki-laki keluar dari rumah	خرج الرجل من البيت	خرج	Fi'il	Fi'il
			الرجل	Isim	Isim
			من	Harf	Harf
			البيت	Isim	Isim
2.	Guru di kelas	المعلم في الفصل	المعلم	Isim	Isim
			في	Harf	Harf
			الفصل	Isim	Isim
3.	Guru menjelaskan pelajaran	يشرح المعلم الدرس	يشرح	Fi'il	Fi'il

No	Kalimat Indonesia	Hasil Terjemahan (Arab)	Token	Label Aktual	Label Prediksi
			المعلم	Isim	Isim
			الدرس	Isim	Isim
4.	Saya pergi ke sekolah	أذهب إلى المدرسة	أذهب	Fi'il	Fi'il
			إلى	Harf	Harf
			المدرسة	Isim	Isim
5.	Seorang anak kecil bermain di lapangan	طفل صغير يلعب في الميدان	طفل	Isim	Isim
			صغير	Isim	Isim
			يلعب	Fi'il	Fi'il
			في	Harf	Harf
			الميدان	Isim	Isim
6.	Dia duduk di kursi	جلس على كرسي	جلس	Fi'il	Fi'il
			على	Harf	Harf
			كرسي	Isim	Isim
7.	Ayahmu bekerja di universitas sebagai dosen	والدك يعمل في الجامعة كمحاضر	والدك	Isim	Isim
			يعمل	Fi'il	Fi'il
			في	Harf	Harf
			الجامعة	Isim	Isim
			ك	Harf	Isim
			محاضر	Isim	Isim

Berdasarkan Tabel 4.14, sistem mampu menerjemahkan kalimat bahasa Indonesia ke dalam bahasa Arab dengan cukup baik sebelum dilakukan proses klasifikasi. Hasil klasifikasi yang diperoleh menunjukkan kesesuaian dengan label aktual pada sebagian besar token, serta konsisten dengan pola prediksi yang dihasilkan pada pengujian menggunakan kalimat bahasa Arab secara langsung.

Berdasarkan hasil pengujian, total token yang dianalisis berjumlah 26 token dengan 25 token berhasil diklasifikasikan sesuai label aktual dan 1 token mengalami kesalahan prediksi. Berdasarkan hasil tersebut, sistem memperoleh tingkat akurasi sebesar:

$$Akurasi = \frac{25}{26} \times 100\% = 96.15\%$$

Hasil ini menunjukkan bahwa sistem mampu mempertahankan performa klasifikasi yang konsisten meskipun input awal diberikan dalam bahasa Indonesia dan terlebih dahulu melalui proses penerjemahan otomatis ke dalam bahasa Arab.

Walaupun demikian, masih terdapat keterbatasan pada token yang mengandung unsur klitik atau imbuhan yang melekat pada kata dasar. Contohnya pada token “والدك” yang mengandung *dhamir muttashil*, serta “كمحاضر” yang mengandung *prefiks harf*, di mana model belum memisahkan komponen tersebut sebagai unit tersendiri. Hal ini menunjukkan bahwa model masih menggunakan pendekatan token utuh tanpa segmentasi morfologis, sehingga beberapa unsur gramatikal belum teridentifikasi secara terpisah.

Temuan tersebut konsisten dengan pola kesalahan pada pengujian menggunakan kalimat bahasa Arab secara langsung. Konsistensi ini menunjukkan bahwa sumber kesalahan utama lebih dipengaruhi oleh keterbatasan model dalam menangani klitik dan segmentasi morfologis dibandingkan proses penerjemahan otomatis.

Pengujian menggunakan input bahasa Indonesia pada penelitian ini dilakukan secara terbatas dan tidak dimaksudkan sebagai evaluasi utama performa model secara statistik. Jumlah kalimat uji yang lebih sedikit dibandingkan pengujian bahasa Arab digunakan untuk merepresentasikan simulasi penggunaan sistem secara nyata (*end-to-end evaluation*), mulai dari proses penerjemahan hingga klasifikasi token. Oleh karena itu, pengujian ini lebih difokuskan untuk mengamati

konsistensi perilaku sistem setelah integrasi seluruh komponen dibandingkan untuk mengukur ulang performa model secara kuantitatif.

4.7.5 Analisis Domain Qur'anic dan *Modern Standard Arabic* (MSA)

Hasil evaluasi menunjukkan bahwa performa model pada tahap implementasi sistem chatbot mengalami perbedaan dibandingkan hasil evaluasi model pada data uji internal. Pada tahap *fine-tuning* dan evaluasi model, model AraBERT memperoleh performa sangat tinggi dengan nilai *F1-score* di atas 0.98 pada data uji internal berbasis dataset MASAQ. Sementara itu, pada tahap implementasi chatbot ditemukan beberapa kesalahan prediksi pada kalimat yang lebih variatif, baik pada input bahasa Arab maupun hasil terjemahan dari bahasa Indonesia.

Perbedaan tersebut tidak dapat dibandingkan secara langsung karena kedua tahap evaluasi menggunakan pendekatan dan tujuan yang berbeda. Evaluasi model dilakukan secara kuantitatif menggunakan dataset berlabel dalam jumlah besar serta metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Sebaliknya, evaluasi implementasi chatbot dilakukan secara terbatas untuk mengamati konsistensi perilaku sistem setelah model diintegrasikan ke dalam aplikasi chatbot.

Pada tahap *fine-tuning*, model dilatih menggunakan dataset *Morphologically and Syntactically Annotated Quran Dataset* (MASAQ) yang berbasis Qur'anic Arabic. Dataset tersebut dipilih karena menyediakan anotasi morfologi bahasa Arab yang lengkap, terstruktur, dan konsisten hingga tingkat token, sehingga sesuai untuk tugas token classification pada penelitian ini. Selain itu, penggunaan dataset publik berlabel memungkinkan proses pelatihan dan evaluasi model dilakukan secara terukur dan replikatif.

Sementara itu, pengujian implementasi chatbot dilakukan menggunakan kalimat *Modern Standard Arabic* (MSA) karena sistem ditujukan untuk mensimulasikan penggunaan nyata pada konteks bahasa Arab modern sehari-hari. Penggunaan MSA pada tahap implementasi bertujuan untuk menguji kemampuan generalisasi model terhadap struktur bahasa di luar domain data pelatihan, sekaligus mengevaluasi konsistensi sistem setelah diintegrasikan ke dalam chatbot.

Secara linguistik, *Qur'anic Arabic* dan MSA memiliki karakteristik yang berbeda. Bahasa Arab dalam Al-Qur'an cenderung memiliki pola sintaksis klasik yang lebih terstruktur dengan distribusi kosakata dan gaya bahasa yang relatif konsisten. Sebaliknya, MSA memiliki variasi struktur kalimat, distribusi kosakata, serta penggunaan klitik yang lebih dinamis sesuai konteks komunikasi modern.

Perbedaan karakteristik tersebut menyebabkan terjadinya *domain shift*, yaitu perbedaan distribusi data antara domain pelatihan dan domain penggunaan sistem. Kondisi ini memengaruhi kemampuan model dalam melakukan generalisasi pada data di luar distribusi pelatihannya, terutama pada token yang mengandung klitik atau struktur morfologis yang lebih kompleks.

Meskipun demikian, hasil pengujian implementasi menunjukkan bahwa model tetap mampu mempertahankan performa klasifikasi yang cukup konsisten pada kalimat MSA. Hal ini menunjukkan bahwa representasi kontekstual yang dipelajari AraBERT selama proses *fine-tuning* masih mampu digunakan untuk mengenali pola kelas kata dasar bahasa Arab pada domain yang berbeda, meskipun masih terdapat keterbatasan pada beberapa kasus morfologis tertentu.

4.7.6 Keterbatasan Penelitian dan Implementasi Sistem

Penelitian ini memiliki beberapa keterbatasan yang perlu ditegaskan agar ruang lingkup dan interpretasi hasil tidak disalahpahami. Keterbatasan tersebut mencakup aspek data, pendekatan pemodelan, serta fungsi sistem chatbot yang dikembangkan.

Pertama, model yang dikembangkan bukan merupakan sistem analisis nahwu bahasa Arab secara lengkap. Klasifikasi yang dilakukan hanya terbatas pada pengenalan jenis kata dasar, yaitu isim, fi'il, dan harf, tanpa mencakup analisis i'rab, fungsi sintaktis, relasi antar kata, maupun struktur gramatikal tingkat lanjut. Oleh karena itu, sistem ini tidak dimaksudkan sebagai pengganti analisis nahwu tradisional.

Kedua, model dilatih menggunakan dataset anotasi morfologi ayat-ayat Al-Qur'an yang memiliki karakteristik linguistik khusus. Meskipun pengujian dilakukan pada kalimat bahasa Arab baku modern (MSA), performa model masih dipengaruhi oleh domain data pelatihan. Hal ini menunjukkan adanya keterbatasan dalam generalisasi lintas domain.

Ketiga, pendekatan yang digunakan adalah token-level classification tanpa segmentasi morfologis klitik secara eksplisit. Akibatnya, token yang mengandung prefiks atau sufiks seperti **و، ل، ب**, maupun dhamir muttashil masih berpotensi tidak terklasifikasi secara terpisah. Keterbatasan ini merupakan konsekuensi dari penggunaan tokenisasi berbasis subword pada model Transformer.

Keempat, sistem chatbot yang dikembangkan tidak dirancang sebagai alat evaluasi pembelajaran. Sistem ini hanya berfungsi untuk menampilkan hasil

klasifikasi kelas kata berdasarkan keluaran model NLP, tanpa melibatkan pengukuran efektivitas pembelajaran, adaptasi terhadap pengguna, maupun evaluasi pedagogis.

Dengan adanya keterbatasan tersebut, penelitian ini difokuskan pada pengembangan model klasifikasi kelas kata serta implementasinya dalam bentuk prototipe chatbot sebagai alat bantu linguistik tahap awal. Penelitian ini juga membuka peluang pengembangan lanjutan, khususnya pada aspek segmentasi morfologis, perluasan domain data, serta integrasi evaluasi pembelajaran di masa mendatang.

4.7.7 Keterkaitan Chatbot dengan Tujuan Penelitian

Pengembangan chatbot dalam penelitian ini merupakan upaya untuk menjawab rumusan masalah mengenai bagaimana mengembangkan dan mengimplementasikan model fine-tuning AraBERT untuk klasifikasi kelas kata bahasa Arab dalam sistem chatbot pembelajaran. Model yang telah melalui tahap pelatihan dan evaluasi kemudian diintegrasikan ke dalam aplikasi, sehingga dapat digunakan untuk melakukan klasifikasi terhadap input teks Arab secara langsung.

Apabila dikaitkan dengan tujuan penelitian, capaian yang diperoleh dapat dilihat dalam beberapa aspek. Pertama, penelitian ini telah mengembangkan model fine-tuning AraBERT untuk melakukan klasifikasi jenis kata (*isim*, *fi'il*, dan *harf*) berdasarkan dataset anotasi morfologi ayat-ayat Al-Qur'an. Hasil evaluasi menunjukkan bahwa model mampu bekerja dengan tingkat konsistensi yang cukup baik, termasuk setelah diintegrasikan ke dalam sistem chatbot.

Kedua, penelitian ini menghasilkan sebuah prototipe chatbot pembelajaran bahasa Arab yang dapat menerima input teks Arab dan menampilkan hasil klasifikasi jenis kata secara otomatis. Proses preprocessing, inferensi model, serta penyajian hasil dapat berjalan dalam satu alur sistem yang terintegrasi, meskipun masih terdapat ruang untuk penyempurnaan lebih lanjut.

Ketiga, chatbot yang dikembangkan berfungsi sebagai alat bantu berbasis NLP untuk menampilkan hasil klasifikasi kelas kata bahasa Arab secara otomatis. Sistem ini berpotensi digunakan sebagai sarana pendukung pembelajaran awal dalam mengenalkan perbedaan antara *isim*, *fi'il*, dan *harf*, tanpa mengklaim efektivitas pedagogis secara langsung.

Dengan mempertimbangkan hasil pengembangan dan implementasi yang telah dilakukan, tujuan penelitian ini dapat dikatakan telah tercapai dalam batasan yang telah ditetapkan. Penelitian ini menghasilkan model klasifikasi berbasis fine-tuning AraBERT sekaligus sebuah prototipe chatbot pembelajaran, yang masih terbuka untuk pengembangan dan penyempurnaan pada penelitian selanjutnya. Karena dalam penelitian ini chatbot difokuskan pada klasifikasi kelas kata dasar (*isim*, *fi'il*, dan *harf*) sebagai tahap awal pembelajaran nahwu, sehingga belum mencakup analisis *i'rab* maupun relasi sintaktis tingkat lanjut.

4.8 Integrasi Penelitian dengan Islam

Penelitian chatbot pembelajaran bahasa Arab berbasis NLP merupakan kajian di bidang kecerdasan buatan dan linguistik komputasional. Selain menghasilkan capaian teknis, penelitian ini juga dapat ditinjau dalam perspektif keilmuan Islam. Keterkaitan ini tidak dimaksudkan sebagai klaim normatif yang luas, melainkan

sebagai kesesuaian arah tujuan penelitian dengan peran bahasa Arab sebagai sarana awal dalam memahami wahyu.

Integrasi tersebut dapat dipahami melalui tiga dimensi hubungan keilmuan dalam Islam, yaitu *ma'allah* (hubungan dengan Allah), *ma'annas* (hubungan dengan manusia), dan *ma'al 'alam* (hubungan dengan lingkungan dan perkembangan ilmu). Ketiga dimensi ini menunjukkan bahwa penelitian tidak secara langsung menghasilkan pemahaman wahyu, tetapi berkontribusi pada tahap awal pembelajaran bahasa Arab, memberikan manfaat dalam proses pendidikan, serta memanfaatkan perkembangan ilmu pengetahuan untuk mendukung pembelajaran yang lebih adaptif.

4.8.1 Dimensi *Ma'allah*

Dimensi *ma'allah* berkaitan dengan hubungan manusia dengan Allah yang berlandaskan tauhid, yaitu keyakinan terhadap keesaan Allah sebagai tujuan utama dalam kehidupan. Dalam Islam, ilmu tidak hanya dipandang sebagai sarana memperoleh pengetahuan, tetapi juga sebagai jalan untuk mengenal Allah dan memperkuat keimanan kepada-Nya (Ibnu Katsir, 2003b). Allah Ta'ala berfirman:

﴿وَمِنَ النَّاسِ وَالْدَّوَابِّ وَالْأَنْعَامِ مُخْتَلِفٌ أَلْوَانُهُ كَذَلِكَ ۚ إِنَّمَا يَخْشَى اللَّهَ مِنْ عِبَادِهِ الْعُلَمَاءُ ۗ إِنَّ اللَّهَ عَزِيزٌ غَفُورٌ﴾

“Dan demikian (pula) di antara manusia, makhluk bergerak yang bernyawa, dan hewan-hewan ternak ada yang bermacam-macam warnanya. Sesungguhnya yang takut kepada Allah di antara hamba-hamba-Nya hanyalah para ulama. Sungguh, Allah Mahaperkasa lagi Maha Pengampun.” (QS. Fatir: 28)

Ayat tersebut menunjukkan bahwa ilmu memiliki peran dalam menumbuhkan rasa takut kepada Allah sebagai bagian dari keimanan dan tauhid. Sejalan dengan itu, kewajiban ini menunjukkan bahwa menuntut ilmu merupakan

sarana untuk mengenal kebesaran Allah dan menguatkan keyakinan terhadap keesaan-Nya (Aniati et al., 2023; Fahrurrosi et al., 2025).

Hal tersebut berkaitan dengan hadis yang diriwayatkan oleh imam Ibnu Majah *hafidzahullah* dari Anas bin Malik *radhiyallahu ‘anhu* Rasulullah ﷺ bersabda:

«طَلَبُ الْعِلْمِ فَرِيضَةٌ عَلَى كُلِّ مُسْلِمٍ»

“Menuntut ilmu itu wajib bagi setiap muslim.” (HR. Ibnu Majah)

Hadis tersebut menunjukkan bahwa proses menuntut ilmu memiliki nilai ibadah dalam Islam. Ketika Allah memerintahkan suatu kewajiban, maka manusia diperintahkan untuk melaksanakannya sebagai bentuk ketaatan kepada-Nya (Khasanah, 2021).

Dalam konteks penelitian ini, pembelajaran bahasa Arab melalui chatbot diposisikan sebagai tahap awal untuk memahami struktur bahasa Al-Qur’an, sehingga mendukung proses memahami makna wahyu secara lebih tepat. Dengan demikian, pengembangan chatbot pembelajaran bahasa Arab tidak hanya berfungsi sebagai inovasi teknologi, tetapi juga sebagai bagian dari upaya menuntut ilmu yang mengarah pada penguatan tauhid.

4.8.2 Dimensi *Ma’annas*

Dimensi *ma’annas* menekankan manfaat penelitian bagi manusia dalam konteks keberagaman bahasa dan komunikasi. Allah Ta’ala berfirman:

﴿يَا أَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِنْ ذَكَرٍ وَأُنْثَىٰ وَجَعَلْنَاكُمْ شُعُوبًا وَقَبَائِلَ لِتَعَارَفُوا ۚ إِنَّ أَكْرَمَكُمْ عِنْدَ اللَّهِ أَتْقَاكُمْ ۚ إِنَّ اللَّهَ عَلِيمٌ خَبِيرٌ ۝﴾

“Wahai manusia! Sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan seorang perempuan, kemudian Kami menjadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah ialah orang yang paling bertakwa. Sungguh, Allah Maha Mengetahui lagi Mahateliti.” (QS. Al-Hujurat: 13)

Ayat ini menunjukkan bahwa manusia diciptakan dalam keberagaman suku dan bahasa, sehingga diperlukan sarana untuk saling memahami. Bahasa Arab sebagai bahasa Al-Qur'an memiliki peran penting dalam menyatukan umat dalam memahami ajaran Islam serta membangun komunikasi yang lebih luas (Ahmed et al., 2022).

Hal ini juga diperkuat dalam hadis yang diriwayatkan oleh Zaid bin Tsabit *radhiyallahu 'anhu*, ketika beliau diperintahkan oleh Rasulullah ﷺ untuk mempelajari bahasa lain demi kebutuhan komunikasi:

«عَنْ زَيْدِ بْنِ ثَابِتٍ قَالَ: أَمَرَنِي رَسُولُ اللَّهِ ﷺ أَنْ أَتَعَلَّمَ لَهُ كَلِمَاتٍ مِنْ كِتَابِ يَهُودَ، وَقَالَ: «إِنِّي وَاللَّهِ مَا آمَنْ يَهُودَ عَلَى كِتَابِي». قَالَ زَيْدٌ: فَمَا مَرَّ بِي نِصْفُ شَهْرٍ حَتَّى تَعَلَّمْتُهُ لَهُ. قَالَ: فَلَمَّا تَعَلَّمْتُهُ كُنْتُ إِذَا كَتَبْتُ إِلَى يَهُودَ كَتَبْتُ إِلَيْهِمْ، وَإِذَا كَتَبُوا إِلَيْهِ قَرَأْتُ لَهُ كِتَابَهُمْ»

“Dari Zaid bin Tsabit radhiyallahu 'anhu beliau berkata: Rasulullah ﷺ memerintahkan aku untuk mempelajari tulisan kaum Yahudi, dan beliau bersabda: “Sesungguhnya aku, demi Allah, tidak merasa aman terhadap tulisan mereka.” Zaid berkata: maka tidak lebih dari setengah bulan aku telah mempelajarinya. Setelah aku mempelajarinya, apabila Rasulullah ﷺ menulis kepada mereka, akulah yang menuliskan surat tersebut, dan apabila mereka menulis kepada beliau, akulah yang membacakan isi tulisan mereka.”

Hadis tersebut menunjukkan bahwa penguasaan bahasa lain merupakan kebutuhan penting dalam komunikasi dan interaksi antar manusia (Khasanah, 2021). Dalam konteks penelitian ini, pengembangan chatbot pembelajaran bahasa Arab memberikan manfaat sebagai media pembelajaran awal yang membantu pengguna dari berbagai latar belakang untuk mengenal struktur dasar bahasa Arab,

khususnya klasifikasi kata (*isim*, *fi'il*, dan *harf*). Dengan demikian, sistem ini mendukung upaya mempermudah pembelajaran bahasa Arab sebagai sarana pemahaman Al-Qur'an dan komunikasi antar manusia.

4.8.3 Dimensi *Ma'al 'Alam*

Dimensi *ma'al 'alam* berkaitan dengan hubungan manusia dengan lingkungan dan perkembangan zaman, khususnya perubahan ekosistem pendidikan akibat teknologi digital. Islam mendorong manusia untuk terus membaca dan belajar, sebagaimana firman Allah Ta'ala:

﴿اَفْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ﴾

“Bacalah dengan nama Tuhanmu yang menciptakan.” (QS. Al-'Alaq: 1)

Perintah membaca ini menunjukkan bahwa proses belajar harus terus berkembang mengikuti perubahan lingkungan kehidupan manusia. Perkembangan teknologi telah mengubah metode pembelajaran dari yang sebelumnya terbatas pada ruang kelas menjadi pembelajaran digital yang fleksibel dan mandiri (Antoun et al., 2020). Penggunaan model NLP berbasis Transformer seperti AraBERT dalam penelitian ini menjadi bentuk pemanfaatan teknologi untuk mendukung pembelajaran bahasa Arab.

Hal ini menegaskan pentingnya integrasi teknologi dalam pendidikan. Chatbot memungkinkan pembelajaran berlangsung secara interaktif dan adaptif, dengan proses yang terstruktur menggunakan data dan metode yang jelas. Dengan demikian, penelitian ini mencerminkan pemanfaatan teknologi dalam mendukung

pembelajaran bahasa Arab sekaligus perkembangan ilmu pengetahuan dalam perspektif Islam.

Selain itu, metode menuntut ilmu dalam Islam juga menekankan pentingnya proses belajar yang terarah dan interaksi antara guru dan peserta didik. Dalam hadis disebutkan:

«خَيْرُكُمْ مَنْ تَعَلَّمَ الْقُرْآنَ وَعَلَّمَهُ»

“Sebaik-baik kalian adalah orang yang belajar Al-Qur'an dan mengajarkannya.”
(HR. Bukhari no. 5027)

Hadis ini menunjukkan pentingnya peran pembelajaran yang terstruktur dan bimbingan dalam proses belajar. Dalam era digital, meskipun teknologi seperti chatbot dan AI mempermudah akses informasi, peran pendidik tetap penting untuk menjaga arah dan kualitas pemahaman (Nisa et al., 2024). Berdasarkan hal tersebut, pemanfaatan teknologi seperti NLP dan chatbot merupakan bentuk integrasi antara perkembangan teknologi modern dengan prinsip pendidikan Islam yang tetap menjaga adab dan tujuan dalam menuntut ilmu.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini berhasil mengembangkan model klasifikasi kelas kata bahasa Arab berbasis *fine-tuning* AraBERT untuk tugas *token classification*. Model yang dilatih menggunakan Dataset *Qur'anic Arabic* (MASAQ) yang telah melalui tahap preprocessing menunjukkan performa yang baik pada data uji internal dengan F1-score sebesar 0,98, yang mengindikasikan kemampuan model dalam mengklasifikasikan token ke dalam kategori *isim*, *fi'il*, dan *harf*.

Model tersebut kemudian berhasil diimplementasikan ke dalam sistem chatbot pembelajaran bahasa Arab berbasis web yang mampu menerima input dalam bahasa Arab maupun bahasa Indonesia melalui proses penerjemahan. Sistem dapat melakukan analisis kelas kata secara otomatis dalam satu alur terintegrasi, sehingga memudahkan pengguna, khususnya pemula, dalam memahami struktur dasar bahasa Arab.

Pada tahap implementasi, performa model mengalami penurunan dibandingkan evaluasi internal akibat perbedaan domain antara data pelatihan berbasis *Qur'anic Arabic* dan data penggunaan berbasis *Modern Standard Arabic* (MSA). Perbedaan ini menimbulkan fenomena *domain shift* yang memengaruhi hasil prediksi. Namun demikian, model tetap menunjukkan kemampuan generalisasi yang cukup baik.

Selain itu, penambahan tahap penerjemahan tidak mengganggu kinerja utama model, melainkan berfungsi sebagai komponen pendukung yang meningkatkan kemudahan penggunaan sistem. Meskipun demikian, model masih memiliki keterbatasan dalam menangani token yang mengandung klitik atau unsur morfologis yang melekat karena belum menggunakan segmentasi morfologis eksplisit. Secara keseluruhan, penelitian ini telah mencapai tujuan dalam mengembangkan model klasifikasi kelas kata serta mengimplementasikannya dalam bentuk chatbot pembelajaran sebagai alat bantu analisis linguistik dasar.

Saran

Pengembangan penelitian selanjutnya dapat diarahkan pada penerapan segmentasi morfologis yang lebih detail agar token yang mengandung klitik dapat dianalisis secara terpisah dan meningkatkan akurasi klasifikasi. Selain itu, perluasan dataset dengan mengintegrasikan corpus *Modern Standard Arabic* (MSA) juga diperlukan untuk mengurangi dampak *domain shift* dan meningkatkan kemampuan generalisasi model.

Dari sisi sistem, pengembangan fitur chatbot dapat dilakukan dengan menambahkan komponen interaktif seperti penjelasan tambahan atau umpan balik adaptif untuk mendukung proses pembelajaran. Evaluasi berbasis pengguna juga perlu dilakukan untuk mengukur efektivitas sistem dalam konteks pembelajaran secara nyata. Selain itu, optimalisasi pada tahap penerjemahan serta eksplorasi model lain dapat dipertimbangkan untuk meningkatkan performa sistem secara keseluruhan.

DAFTAR PUSTAKA

- Abdul Hamid, M. M. (2019). *Al-Tuhfah al-Saniyah Syarah al-Muqaddimah al-Ajurumiyyah* (M. Ridwan, Trans.). Perniagaan Jahabersa. https://www.researchgate.net/publication/334316141_TERJEMAH_SYA_RAH_JURUMIYAH_AL-TUHFAH_AL-SANIYAH_Syarah_al-Muqaddimah_al-Ajurumiyyah
- Abo-Elghit, A. H., Hamza, T., & Al-Zoghby, A. (2022). Embedding Extraction for Arabic Text Using the AraBERT Model. *Computers, Materials & Continua*, 72(1), 1967–1994. <https://doi.org/10.32604/cmc.2022.025353>
- Ahmed, A., Ali, N., Alzubaidi, M., Zaghouani, W., Abd-alrazaq, A., & Househ, M. (2022). Arabic chatbot technologies: A scoping review. *Computer Methods and Programs in Biomedicine Update*, 2, 100057. <https://doi.org/10.1016/j.cmpbup.2022.100057>
- Alazzam, B. A., Alkhatib, M., & Shaalan, K. (2023). Arabic Educational Neural Network Chatbot. *Information Sciences Letters*, 12(6), 2579–2589. <https://doi.org/10.18576/isl/120654>
- Al-Hajj, M., Jarrar, M., & Birzeit University Birzeit, Palestine. (2021). ArabGlossBERT: Fine-Tuning BERT on Context-Gloss Pairs for WSD. *Proceedings of the Conference Recent Advances in Natural Language Processing - Deep Learning for Natural Language Processing Methods and Applications*, 35–43. https://doi.org/10.26615/978-954-452-072-4_005
- Al-harbi, N. K., & Algheith, M. (2025). Fine-Tuning Arabic and Multilingual BERT Models for Crime Classification to Support Law Enforcement and Crime Prevention. *International Journal of Advanced Computer Science and Applications* (IJACSA), 16. https://thesai.org/Downloads/Volume16No5/Paper_44-Fine_Tuning_Arabic_and_Multilingual_BERT_Models.pdf
- Alruqi, T. N., & Alzahrani, S. M. (2023). Evaluation of an Arabic Chatbot Based on Extractive Question-Answering Transfer Learning and Language Transformers. *AI*, 4(3), 667–691. <https://doi.org/10.3390/ai4030035>
- Al-Sheikh, I., & Mohd, M. (2019). A Quranic Dataset for Text Recognition. *Proceedings of the Proceedings of the 1st International Conference on Informatics, Engineering, Science and Technology, INCITEST 2019, 18 July 2019, Bandung, Indonesia*. Proceedings of the 1st International Conference on Informatics, Engineering, Science and Technology,

INCITEST 2019, 18 July 2019, Bandung, Indonesia.
<https://doi.org/10.4108/eai.18-7-2019.2287842>

Aniati, I., Dalimunthe, I. S., & Hayati, N. (2023). The urgency of arabic in the study of the Qur'an. *At Turots: Jurnal Pendidikan Islam*, 646–654.
<https://doi.org/10.51468/jpi.v5i1.256>

Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT: Transformer-based Model for Arabic Language Understanding. *ACL Antology*.
<https://aclanthology.org/2020.osact-1.2/>

Badawi, A. (2021). THE EFFECTIVENESS OF NATURAL LANGUAGE PROCESSING (NLP) AS A PROCESSING SOLUTION AND SEMANTIC IMPROVEMENT. *International Journal of Economic, Technology and Social Sciences (Injects)*, 2(1), 36–44.
<https://doi.org/10.53695/injects.v2i1.194>

El Moubtahij, H., Abdelali, H., & Tazi, E. B. (2022). AraBERT transformer model for Arabic comments and reviews analysis. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 11(1), 379.
<https://doi.org/10.11591/ijai.v11.i1.pp379-387>

El-Alami, F., Ouatik El Alaoui, S., & En Nahnahi, N. (2022). Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization. *Journal of King Saud University - Computer and Information Sciences*, 34(10), 8422–8428.
<https://doi.org/10.1016/j.jksuci.2021.02.005>

Fahrurrosi, L., Siregar, M. T. I., Az-Zuhdy, A. H., & Hesim Muzedi. (2025). Analisis Kewajiban Menuntut Ilmu Dalam Islam Perpspektif Al-Qur'an Dan Hadis. *Jurnal IHSAN Jurnal Pendidikan Islam*, 3(1), 348–357.
<https://doi.org/10.61104/ihsan.v3i1.834>

Faroid, U., Maulidyaningrum, H., Hadi, W., & Salsabila, U. H. (2025). Chatbot Mondly Sebagai Media Inovatif Dalam Pembelajaran Bahasa Arab. *Jurnal Ilmiah Pendidikan Citra Bakti*, 12(1), 253–264.
<https://doi.org/10.38048/jipcb.v12i1.5093>

Gong, J., Ma, H., Teng, Z., Teng, Q., Zhang, H., Du, L., Chen, S., Bhuiyan, M. Z. A., Li, J., & Liu, M. (2020). Hierarchical Graph Transformer-Based Deep Learning Model for Large-Scale Multi-Label Text Classification. *IEEE Access*, 8, 30885–30896. <https://doi.org/10.1109/ACCESS.2020.2972751>

Hanif, M. A. H. (2025). Pembelajaran Bahasa Arab Sebagai Stimulus Bagi Santri Tahfizh Al-Quran di Pp. Al-Munawwir Komplek Ribathul Quran II Untuk Memahami Makna Ayat Al-Quran. *Raudhah Proud To Be Professionals :*

- Ibnu Katsir, I. bin U. (2003a). *Lubaabut Tafsir Min Ibni Katsir (Tafsir Ibnu Katsir) Jilid 4* (M. A. E. M. Ghoffar & A. I. Al-Atsari, Trans.; Vol. 4). Pustaka Imam asy-Syafi'i. <https://ajibhanani.blogspot.com/2017/07/kitab-kitab.html>
- Ibnu Katsir, I. bin U. (2003b). *Lubaabut Tafsir Min Ibni Katsir (Tafsir Ibnu Katsir) Jilid 6* (M. A. E. M. Ghoffar & A. I. Al-Atsari, Trans.; Vol. 6). Pustaka Imam asy-Syafi'i. <https://ajibhanani.blogspot.com/2017/07/kitab-kitab.html>
- Indu, P. V., & K, V. (2020). Research designs-an overview. *Kerala Journal of Psychiatry*, 32(1). <https://doi.org/10.30834/KJP.32.1.2019.179>
- Jafari, A. (2022). *Comparison Study Between Token Classification and Sequence Classification In Text Classification* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2211.13899>
- Khasanah, W. (2021). Kewajiban Menuntut Ilmu dalam Islam. *Jurnal Riset Agama*, 1(2), 296–307. <https://doi.org/10.15575/jra.v1i2.14568>
- Ma, X., & Hovy, E. (2016). End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1064–1074. <https://doi.org/10.18653/v1/P16-1101>
- Malik, I. (2006). شرح ألفية ابن مالك (A. L. bin M. Al-Khatib, Ed.). <https://archive.org/details/AlfiahIbnMalikUthaimin1> (Original work published Dar Al-'Urubah)
- Matrane, Y., Benabbou, F., & Ellaky, Z. (2024). Enhancing Moroccan Dialect Sentiment Analysis Through Optimized Preprocessing and Transfer Learning Techniques. *IEEE Access*, 12, 187756–187777. <https://doi.org/10.1109/ACCESS.2024.3514934>
- Moradi, M., Rahmanimanesh, M., & Shahzadi, A. (2022). On Evaluating the Collaborative Research Areas: A Case Study. *Journal of King Saud University - Computer and Information Sciences*, 34(2), 408–420. <https://doi.org/10.1016/j.jksuci.2019.11.005>
- Mulyadi, M. (2013). Riset Desain Dalam Metodologi Penelitian. *Jurnal Studi Komunikasi Dan Media*, 16(1), 71. <https://doi.org/10.31445/jskm.2012.160106>

- Nguyen, N., & Guo, Y. (2007). Comparisons of sequence labeling algorithms and extensions. *Proceedings of the 24th International Conference on Machine Learning*, 681–688. <https://doi.org/10.1145/1273496.1273582>
- Nisa, S. C., Novita Pamujianti, A., Agung Purnomo, A., & Abbas, N. (2024). Etika dan Metode Menuntut Ilmu Perspektif Hadits Nabi Muhammad. *ISEDU : Islamic Education Journal*, 2(2), 6–16. <https://doi.org/10.59966/isedu.v2i2.1229>
- Pouramini, A., & Faili, H. (2024). *Matching Pre-Trained Language Models with Specific Tasks: Fine-Tuning and Prompt-Tuning Strategies*. SSRN. <https://doi.org/10.2139/ssrn.4702919>
- Puvis De Chavannes, L. H., Kongsbak, M. G. K., Rantza, T., & Derczynski, L. (2021). Hyperparameter Power Impact in Transformer Language Model Training. *Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing*, 96–118. <https://doi.org/10.18653/v1/2021.sustainlp-1.12>
- Rini, R. (2019). Ushul al-Nahwi al-Arabi: Kajian Tentang Landasan Ilmu Nahwu. *Arabiyatuna : Jurnal Bahasa Arab*, 3(1), 145. <https://doi.org/10.29240/jba.v3i1.773>
- Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A Primer in BERTology: What We Know About How BERT Works. *Transactions of the Association for Computational Linguistics*, 8, 842–866. https://doi.org/10.1162/tac1_a_00349
- Roziqi, A. K., & Bakar, M. Y. A. (2025). Epistemologi Ilmu Nahwu: Studi Ilmu Tata Bahasa Dalam Perspektif Filsafat Ilmu. *Al-Fakkar*, 6(1), 56–75. <https://doi.org/10.52166/alf.v6i1.7983>
- Sawalha, M. (2024). *MASQA: Morphologically and Syntactically-Annotated Quran Dataset* [Dataset]. Mendeley Data. <https://doi.org/10.17632/9YVRZXTMR.2>
- Shohib, Muh. I., & Fuadah, N. (2022). Ushul An-Nahwi (Kajian Epistimologi Sintaksis Arab). *Lisan An Nathiq : Jurnal Bahasa Dan Pendidikan Bahasa Arab*, 4(1), 83–94. <https://doi.org/10.53515/lan.v4i1.4934>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (Version 7). arXiv. <https://doi.org/10.48550/ARXIV.1706.03762>
- Yacoub, R., & Axman, D. (2020). Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models.

Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, 79–91. <https://doi.org/10.18653/v1/2020.eval4nlp-1.9>

Zaid, L. N., Utami, W. F., Fauzi, M. R., & Nasution, S. (2024). Peran Linguistik dalam Mempelajari Struktur Bahasa Arab. *Semantik : Jurnal Riset Ilmu Pendidikan, Bahasa Dan Budaya*, 3(1), 11–19. <https://doi.org/10.61132/semantik.v3i1.1320>

Zeroual, I., Lakhouaja, A., & Belahbib, R. (2017). Towards a standard Part of Speech tagset for the Arabic language. *Journal of King Saud University - Computer and Information Sciences*, 29(2), 171–178. <https://doi.org/10.1016/j.jksuci.2017.01.006>