BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan ekosistem digital dalam satu dekade terakhir telah mempercepat distribusi informasi lintas negara secara *real-time* melalui media sosial, portal berita *online*, dan platform berbagi konten. Namun percepatan arus informasi tersebut juga diikuti oleh peningkatan signifikan dalam penyebaran hoaks dan misinformasi yang beredar dalam berbagai bahasa. Studi menunjukkan bahwa sekitar 33% klaim misinformasi yang berulang tidak terbatas pada satu rumpun bahasa saja, melainkan aktif berpindah mengikuti pola komunikasi lintas bangsa [1]. Fenomena ini diperparah oleh kemudahan translasi otomatis dan adaptasi ulang konten yang memungkinkan sebuah narasi hoaks beredar dalam beberapa bahasa hanya dalam waktu singkat [2]. Misinformasi sering kali berkembang dan beradaptasi lintas batas bahasa, termasuk pola adopsi konten berbahasa Inggris oleh pengguna bahasa lain dalam konteks pandemi COVID-19 [3]. Dalam konteks Asia Tenggara, bahasa Indonesia/Melayu dan Inggris membentuk ekosistem multibahasa yang saling terhubung dalam penyebaran berita dan percakapan *online*, termasuk konten yang belum terverifikasi. Di Malaysia, analisis 108.246 tweet terkait COVID-19 menunjukkan sekitar 67% berbahasa Melayu dan 27% berbahasa Inggris, dengan banyak cuitan yang mencampur beberapa bahasa sekaligus [4]. Kondisi serupa juga ditemukan di Indonesia, di mana bahasa Indonesia dan Inggris digunakan secara bergantian maupun bersamaan sebagai medium penyebaran hoaks di platform seperti Facebook, WhatsApp, Instagram, dan Twitter/X, dengan konten yang kerap memanfaatkan bahasa nonbaku untuk memengaruhi emosi serta persepsi publik [5]. Realitas multibahasa ini menegaskan bahwa pendekatan deteksi hoaks berbasis satu bahasa tidak lagi memadai, sehingga diperlukan sistem yang mampu mengenali dan menganalisis konten lintas bahasa secara terintegrasi untuk menghadapi dinamika penyebaran misinformasi di kawasan Asia Tenggara.

Perkembangan ekosistem digital dalam satu dekade terakhir telah mempercepat distribusi informasi lintas negara secara *real-time* melalui media

sosial, portal berita *online*, dan platform berbagi konten. Percepatan arus informasi tersebut juga diikuti oleh peningkatan signifikan dalam penyebaran hoaks dan misinformasi yang beredar dalam berbagai bahasa. Studi menunjukkan bahwa sekitar 33% klaim misinformasi yang berulang tidak terbatas pada satu rumpun bahasa saja, melainkan aktif berpindah mengikuti pola komunikasi lintas bangsa [1]. Fenomena ini diperparah oleh kemudahan translasi otomatis dan adaptasi ulang konten yang memungkinkan sebuah narasi hoaks beredar dalam beberapa bahasa hanya dalam waktu singkat [2]. Misinformasi sering kali berkembang dan beradaptasi lintas batas bahasa, termasuk pola adopsi konten berbahasa Inggris oleh pengguna bahasa lain dalam konteks pandemi COVID-19 [3].

Merespons tantangan penyebaran hoaks multibahasa yang semakin kompleks di kawasan Asia Tenggara, deteksi berita palsu atau hoaks menjadi salah satu tugas paling krusial dalam *Natural Language Processing* (NLP), khususnya pada bidang klasifikasi teks lintas bahasa. Dalam konteks ini, perkembangan arsitektur *deep learning*, terutama model berbasis *transformer*, menjadi semakin relevan karena kemampuannya memahami konteks dan variasi bahasa secara lebih mendalam, sehingga lebih efektif untuk menangani dinamika teks multibahasa dibandingkan *machine learning* atau *deep learning konvensional*. Keunggulan ini terutama berasal dari mekanisme *self-attention* yang memungkinkan model memahami hubungan antar token secara kontekstual, berbagai studi dalam ACL dan EMNLP menunjukkan bahwa model *transformer* unggul secara signifikan dalam tugas deteksi misinformasi dibandingkan metode berbasis fitur tradisional seperti SVM atau Naïve Bayes [6][7]. Dalam skenario *multilingual*, kebutuhan akan representasi lintas bahasa semakin penting agar model mampu melakukan *transfer learning* antar bahasa tanpa pelatihan terpisah, dan superioritas model transformer dalam hal ini telah dikonfirmasi secara konsisten oleh sejumlah studi komparatif.

Studi terbaru dalam klasifikasi teks lintas bahasa menunjukkan bahwa pemanfaatan model berbasis *transformer* secara signifikan melampaui pendekatan tradisional dalam mendeteksi misinformasi dan hoaks. Penelitian yang mengeksplorasi deteksi misinformasi multibahasa menemukan bahwa mBERT dan XLM-RoBERTa mampu mencapai tingkat akurasi masing-masing sebesar 92% dan 91,41% dalam menangkap nuansa kontekstual bahasa [8]. Pencapaian ini

menunjukkan lonjakan performa yang tajam jika dibandingkan dengan model *baseline* berarsitektur *deep learning* konvensional seperti Convolutional Neural Network (CNN) dan Bi-directional Gated Recurrent Unit (BiGRU) yang hanya bertahan pada tingkat akurasi 81,5% dan 83,3% [8]. Dominasi arsitektur *transformer* juga dikonfirmasi oleh studi lain, di mana model seperti XLM-RoBERTa sukses mencatatkan akurasi 83% pada *dataset* teks bias yang sangat kompleks, jauh mengungguli metode *machine learning* klasik seperti Support Vector Machine (SVM) yang kemampuannya sering kali merosot hingga di bawah 75% saat dihadapkan pada skenario *cross-lingual*. Bahkan pada skenario bahasa bersumber daya sangat rendah seperti Kinyarwanda-Kirundi, mBERT (88,6%) tetap kompetitif melampaui BiGRU (83,3%), CNN (81,5%), dan char-CNN (79,8%), meski sedikit di atas AfriBERT (88,3%) yang dirancang khusus untuk bahasa Afrika [9]. Berbagai temuan ini menegaskan bahwa keluarga model *transformer*, khususnya mBERT, DistilmBERT, dan XLM-RoBERTa, merupakan fondasi yang sangat kuat untuk dikembangkan dalam sistem deteksi hoaks yang beroperasi pada berbagai bahasa secara bersamaan.

Arsitektur *transformer* telah merevolusi bidang NLP melalui mekanisme *self-attention* yang memungkinkan pemahaman konteks secara dua arah dan penangkapan hubungan antar kata secara mendalam [10]. Salah satu implementasi paling berpengaruh dalam klasifikasi lintas bahasa adalah Multilingual BERT (mBERT), yang dilatih menggunakan teknik *Masked Language Modeling* (MLM) pada korpus Wikipedia mencakup 104 bahasa secara simultan [11]. Keunggulan mBERT terletak pada kemampuannya melakukan transfer pengetahuan antar bahasa secara efisien melalui representasi semantik bersama tanpa memerlukan pelatihan ulang secara spesifik untuk setiap bahasa target [12]. Untuk mengatasi keterbatasan pada bahasa bersumber daya rendah, XLM-RoBERTa juga dapat dipilih karena dilatih pada *dataset* CommonCrawl yang jauh lebih masif dibandingkan korpus Wikipedia [13]. XLM-RoBERTa mengoptimalkan prosedur *pre-training* dengan menghilangkan tugas *Next Sentence Prediction* (NSP) serta menggunakan ukuran *batch* yang lebih besar, sehingga menghasilkan kapasitas generalisasi lintas bahasa yang lebih superior untuk mendeteksi hoaks yang manipulatif [13]. Secara arsitektur, baik mBERT maupun XLM-RoBERTa

memiliki jutaan parameter yang mampu menangkap fitur linguistik kompleks yang sering kali luput dari metode tradisional seperti TF-IDF atau *word embedding* statis [14]. Konsistensi performa dari model-model *transformer* ini menjadikannya standar utama dalam tugas klasifikasi teks multibahasa, meskipun tantangan terkait kebutuhan sumber daya komputasi yang tinggi tetap menjadi fokus utama dalam pengembangan sistem yang efisien [15].

Sebagai alternatif yang lebih ringan, DistilmBERT dikembangkan melalui pendekatan *knowledge distillation*, yaitu *transfer* pengetahuan dari model besar (*teacher*) ke model lebih kecil (*student*). DistilBERT mampu memangkas ukuran model sebesar 40% dan berjalan 60% lebih cepat, sambil mempertahankan 97% kemampuan pemahaman bahasa BERT asli, dengan hanya 66 juta parameter dan enam layer *transformer* [16]. Versi multilingualnya, DistilmBERT, dilatih pada 104 bahasa dengan waktu inferensi rata-rata dua kali lebih cepat dibandingkan mBERT [17]. Studi komparatif menunjukkan waktu *fine-tuning* yang 21,5% - 66,9% lebih singkat dibandingkan model penuh [18]. Namun kemampuan *transfer* lintas bahasa pada model *distilled* belum sepenuhnya sekuat mBERT. Perbandingan pada klasifikasi teks menunjukkan bahwa performa model *distilled* seringkali mencatatkan akurasi dan *F1-score* yang sedikit lebih rendah dibandingkan mBERT dan XLM-RoBERTa pada kondisi yang sama [19].

Meskipun *full fine-tuning* telah terbukti efektif dalam mengadaptasi model *transformer* untuk tugas deteksi hoaks, pendekatan ini menuntut pembaruan seluruh parameter model secara simultan, yang pada arsitektur seperti XLM-RoBERTa dengan 278 juta parameter mengakibatkan konsumsi memori GPU dan biaya komputasi yang sangat tinggi [20]. Pada skenario *few-shot cross-lingual*, di mana data bahasa target hanya tersedia dalam jumlah sangat terbatas, *full fine-tuning* justru berisiko mengganggu representasi lintas bahasa yang terbentuk selama *pre-training* karena pembaruan massal terhadap parameter model menyebabkan fenomena yang dikenal sebagai *catastrophic forgetting* [21]. Sebagai alternatif, *Parameter-Efficient Fine-Tuning* (PEFT) berbasis *Low-Rank Adaptation* (LoRA) menawarkan solusi dengan membekukan bobot *pre-trained* dan menyisipkan pasangan matriks berperingkat rendah yang hanya memperbarui kurang dari 1% parameter total model [22]. Studi komparatif menunjukkan bahwa LoRA mampu

mempertahankan performa yang baik terhadap *full fine-tuning* dengan pengurangan *trainable parameters* hingga 140 kali lipat pada berbagai tugas klasifikasi teks multilingual [23].

Namun demikian, LoRA standar menerapkan rank yang seragam di seluruh lapisan model, padahal setiap kelompok lapisan berkontribusi secara berbeda terhadap performa tugas, dengan lapisan atas (*top layers*) lebih bertanggung jawab terhadap adaptasi *task-specific* sementara lapisan bawah (*bottom layers*) menyimpan representasi linguistik universal [24]. Keterbatasan *rank* seragam ini mendorong dikembangkannya Adaptive-Rank LoRA (ARL-LoRA), yaitu varian LoRA yang mengalokasikan kapasitas *rank* secara berbeda pada setiap kelompok lapisan sesuai dengan peran fungsionalnya: *rank* kecil (r=8) pada lapisan bawah untuk menjaga kestabilan representasi lintas bahasa yang telah terbentuk selama *pre-training*, *rank* sedang (r=16) pada lapisan tengah untuk memfasilitasi transfer semantik, dan *rank* besar (r=64) pada lapisan atas untuk memaksimalkan kapasitas adaptasi *task-specific* [25]. Strategi alokasi *rank* adaptif pada ARL-LoRA ini terbukti menghasilkan adaptasi yang lebih optimal dibandingkan LoRA standar, khususnya pada tugas yang memerlukan diferensiasi kapasitas representasi antar kelompok lapisan [25]. Pada tugas *cross-lingual* dengan data target minimal, ARL-LoRA masih menghadapi tantangan tambahan karena beroperasi pada level *transformer layer* dan belum menyentuh dua komponen kritis, yaitu *bias terms* yang mengatur sensitivitas keputusan setiap neuron dan *token embedding* yang menentukan posisi representasi kata-kata bahasa target dalam ruang semantik model [9].

BitFit, sebuah metode PEFT yang hanya melatih *bias terms* model, terbukti mampu menyamai performa *full fine-tuning* pada data berukuran kecil hingga menengah meski hanya mengubah sekitar 0,09% parameter [26]. Pemilihan BitFit sebagai komponen pelengkap ARL-LoRA didasarkan pada kesenjangan fungsional yang belum tertangani oleh ARL-LoRA secara mandiri: ARL-LoRA beroperasi pada level *transformer layer* melalui matriks adaptasi berperingkat rendah yang memodelkan rotasi dan *scaling* dalam ruang representasi, namun tidak menyentuh *bias terms* yang mengatur sensitivitas keputusan setiap neuron [9][25]. Secara geometris, BitFit bersifat komplementer terhadap LoRA karena BitFit memodelkan

translasi sementara LoRA memodelkan rotasi, sehingga kombinasi keduanya mencakup lebih banyak jenis transformasi yang dibutuhkan dalam adaptasi lintas bahasa [26]. Dibandingkan metode PEFT alternatif seperti Adapter yang menyisipkan modul *bottleneck* baru di antara lapisan *transformer*, BitFit tidak menambahkan parameter baru pada arsitektur sehingga tidak meningkatkan latensi inferensi; sementara IA3 yang bekerja melalui *rescaling* aktivasi secara umum kurang efektif pada skenario data berlabel sangat terbatas dibandingkan modifikasi *bias terms* yang lebih langsung memengaruhi distribusi keputusan model [20][23].

Meskipun kombinasi ARL-LoRA dan BitFit meningkatkan kapasitas adaptasi model secara signifikan, keduanya beroperasi pada level *transformer layer* dan belum mengatasi permasalahan yang berasal dari level *input representation*, yaitu posisi vektor *embedding* untuk *token* bahasa target yang bersumber daya rendah [27]. Studi terkini menunjukkan bahwa *embedding layer* dan lapisan yang berdekatan dengannya menyimpan representasi paling spesifik terhadap bahasa, sehingga adaptasi pada level ini menjadi krusial untuk meningkatkan kualitas *cross-lingual transfer* [28]. *Selective Embedding Adaptation* (SEA) menjawab kebutuhan tersebut dengan hanya mengaktifkan pembaruan gradien pada baris *embedding* untuk *token* yang muncul dalam data bahasa target, sehingga representasi *token* bahasa target dapat dioptimalkan tanpa memerlukan pelatihan ulang terhadap seluruh matriks *embedding* yang berukuran sangat besar [29]. Pendekatan ini dipilih dibandingkan *prefix tuning* yang menyisipkan vektor konteks artifisial di awal sekuens, karena *prefix tuning* tidak secara langsung mengoptimalkan representasi *token* asli dari bahasa target melainkan hanya menambahkan sinyal kontekstual eksternal yang bergantung pada kemampuan model menginterpolasikan bahasa asing dari vektor prefiks tersebut [28]. Pada bahasa dengan *token* yang jarang muncul dalam korpus *pre-training* seperti Bahasa Melayu, pendekatan langsung melalui pembaruan selektif vektor *embedding* terbukti lebih efektif karena secara eksplisit memperbaiki representasi awal yang kurang optimal akibat minimnya paparan bahasa tersebut selama *pre-training* [28][29]. Integrasi ketiga komponen, yaitu ARL-LoRA untuk adaptasi *transformer layer* dengan rank berbeda per kelompok lapisan, BitFit untuk fleksibilitas *bias*, dan SEA untuk adaptasi *embedding* selektif, secara teoretis membentuk kerangka PEFT *hybrid* yang

mengintervensi model pada tiga level yang berbeda, yakni *input representation*, *transformation*, dan *output decision*, sehingga diharapkan mampu menutup kesenjangan performa terhadap *full fine-tuning* dengan jumlah *trainable parameter* yang jauh lebih kecil [22][25][26].

Bahasa Indonesia ditetapkan sebagai bahasa sumber pelatihan karena dua pertimbangan utama. Pertama, ekosistem berita daring berbahasa Indonesia tergolong besar dan aktif, dengan jutaan konten yang diproduksi dan dikonsumsi setiap harinya melalui portal berita nasional maupun media sosial [5]. Kedua, *dataset* berlabel berkualitas yang dapat diakses publik sudah tersedia dalam jumlah yang memadai, mencakup arsip berita dari CNN Indonesia, Kompas, dan Tempo untuk kelas faktual, serta basis data hoaks dari TurnBackHoax yang telah diverifikasi secara editorial [5]. Bahasa Melayu ditempatkan sebagai bahasa target *few-shot* karena posisinya yang unik dalam ekosistem linguistik Asia Tenggara. Kedekatan tipologis, morfologis, dan leksikal antara Bahasa Indonesia dan Melayu menjadikan keduanya sebagai pasangan bahasa yang ideal untuk menguji hipotesis bahwa representasi lintas bahasa yang terbentuk selama pre-training pada korpus multibahasa dapat dimanfaatkan secara efektif pada bahasa serumpun meskipun data berlabelnya sangat terbatas [10]. Pasangan bahasa ini juga memiliki relevansi praktis yang tinggi, mengingat penyebaran hoaks di kawasan Asia Tenggara kerap melintasi batas linguistik antara komunitas berbahasa Indonesia dan Melayu [4].

LIME dipilih sebagai metode *Explainable* AI atas pertimbangan metodologis dan praktis yang spesifik untuk konteks penelitian ini. Sejumlah metode XAI telah diusulkan dalam literatur, termasuk SHAP (SHapley Additive exPlanations), *attention visualization*, dan *gradient-based attribution* seperti Integrated Gradients. Keunggulan utama LIME terletak pada sifatnya yang *model-agnostic*, yang memungkinkan penerapan konsisten pada ketiga arsitektur berbeda (XLM-RoBERTa, mBERT, DistilmBERT) tanpa memerlukan modifikasi internal pada masing-masing model [30]. Dalam penelitian komparatif, konsistensi mekanisme ekstraksi penjelasan ini sangat krusial karena menjamin bahwa perbedaan pola interpretabilitas yang teridentifikasi benar-benar mencerminkan perbedaan strategi representasi antar model, bukan artefak metodologis [31]. *Attention visualization* tidak digunakan karena literatur terkini menunjukkan bahwa bobot *attention* tidak

selalu berkorelasi langsung dengan kontribusi fitur terhadap prediksi, sehingga interpretasinya sebagai penjelasan kausal bersifat problematis [32]. SHAP tidak digunakan sebagai metode utama karena kompleksitas komputasionalnya yang jauh lebih tinggi pada model berbasis *transformer* dengan ribuan fitur teks, yang secara praktis membatasi jumlah sampel yang dapat dianalisis dalam batas waktu yang wajar. Beberapa studi terdahulu yang menggunakan LIME pada sistem deteksi hoaks berbasis *transformer* juga mengonfirmasi bahwa LIME mampu mengidentifikasi fitur linguistik yang relevan secara semantik dan konsisten dengan penilaian ahli bahasa [31][32].

Penelitian ini mengusulkan ARL-LoRA^{+}, sebuah kerangka *hybrid Parameter-Efficient Fine-Tuning* yang mengintegrasikan tiga mekanisme komplementer untuk mengatasi keterbatasan LoRA standar pada skenario *few-shot cross-lingual hoax detection*. Komponen pertama, *Adaptive-Rank LoRA* (ARL-LoRA), menerapkan rank yang berbeda pada setiap kelompok lapisan *transformer* dengan alokasi rank kecil pada lapisan bawah yang menyimpan representasi universal, rank sedang pada lapisan tengah yang bertanggung jawab atas *cross-lingual transfer*, dan rank besar pada lapisan atas yang melakukan adaptasi *task-specific* [25]. Komponen kedua, *Bias Tuning* (BitFit), mengaktifkan seluruh *bias terms* di setiap lapisan untuk memberikan fleksibilitas pergeseran keputusan yang melengkapi kapasitas rotasi dari matriks LoRA [26]. Komponen ketiga, *Selective Embedding Adaptation* (SEA), secara selektif memperbarui vektor *embedding* hanya untuk token yang terdeteksi muncul dalam data Bahasa Melayu melalui mekanisme *gradient masking*, sehingga representasi token bahasa target dapat dioptimalkan tanpa mengganggu ratusan ribu token lainnya yang sudah terposisi dengan baik dalam ruang semantik model [28][29]. Kerangka ARL-LoRA^{+} ini divalidasi melalui eksperimen komparatif pada tiga arsitektur *multilingual transformer* yaitu Multilingual BERT, XLM-RoBERTa, dan DistilmBERT, dengan skema pelatihan pada Bahasa Indonesia dan Inggris serta evaluasi *few-shot* (100 sampel per kelas) pada Bahasa Melayu sebagai bahasa target. Perbandingan dilakukan antara skenario *full fine-tuning* dan skenario ARL-LoRA^{+}, dilengkapi analisis *Explainable AI* berbasis LIME untuk mengidentifikasi kontribusi fitur leksikal terhadap keputusan klasifikasi pada setiap arsitektur dan bahasa [31]. Penelitian ini diharapkan

menghasilkan bukti konkret mengenai efektivitas pendekatan PEFT hybrid dalam mengatasi kesenjangan performa terhadap *full fine-tuning* pada skenario *few-shot cross-lingual*, sekaligus memberikan implikasi praktis bagi pengembangan sistem deteksi hoaks multilingual yang efisien, transparan, dan dapat dijalankan dengan sumber daya komputasi terbatas di kawasan Asia Tenggara.

1.2 Rumusan Masalah

Rumusan masalah yang ada dalam penelitian ini, yaitu:

1. Bagaimana performa ketiga model *multilingual transformer* (XLM-RoBERTa, mBERT, dan DistilmBERT) dalam mendeteksi hoaks pada skenario *cross-lingual few-shot* ke Bahasa Melayu, dan apakah metode ARL-LoRA^{+} mampu meningkatkan F1-*weighted* Bahasa Melayu dibandingkan *full fine-tuning* konvensional pada masing-masing arsitektur model?
2. Bagaimana profil efisiensi komputasi ARL-LoRA^{+} dibandingkan *full fine-tuning* dalam hal jumlah *trainable parameter*, waktu pelatihan, dan penggunaan memori GPU puncak, serta bagaimana kontribusi setiap komponen (ARL, BitFit, dan SEA) terhadap total peningkatan kinerja berdasarkan hasil *ablation study*?
3. Bagaimana stabilitas kinerja ARL-LoRA^{+} lintas berbagai kondisi inisialisasi acak (lima *random seed*) dan variasi *sampling few-shot* Bahasa Melayu (mode *fix* dan *resample*), serta faktor arsitektural apa yang menentukan apakah suatu model dapat memperoleh manfaat dari penerapan ARL-LoRA^{+}?
4. Pola linguistik apa yang diidentifikasi oleh metode *Explainable AI* berbasis LIME (*Local Interpretable Model-Agnostic Explanations*) sebagai penanda hoaks dan berita real pada ketiga bahasa, dan bagaimana perbedaan strategi pengambilan keputusan antara ARL-LoRA^{+} dan *full fine-tuning* lintas XLM-RoBERTa, mBERT, dan DistilmBERT?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini diantaranya:

1. Konfigurasi *few-shot* pada Bahasa Melayu ditetapkan sebesar 100 sampel per kelas dan bersifat tetap; penelitian ini tidak membandingkan berbagai ukuran *shot* lainnya.

2. Periode pengambilan data dibatasi pada berita *online* yang dipublikasikan dalam rentang waktu 2021 hingga 2025.
3. Modalitas data terbatas pada teks berita tanpa melibatkan informasi tambahan seperti gambar, video, *metadata* media sosial, atau sumber penerbit sebagai variabel terpisah.
4. Klasifikasi dibatasi pada dua kelas yaitu *hoax* dan *real*, tanpa kategori tambahan seperti *satire* atau tingkat kredibilitas bertingkat.
5. Model *transformer* yang digunakan dibatasi pada tiga arsitektur *multilingual transformer*, yaitu Multilingual BERT (mBERT), DistilmBERT, dan XLM-RoBERTa.
6. Metode PEFT yang diterapkan adalah ARL-LoRA^{+} yang terdiri dari *Adaptive-Rank LoRA*, BitFit, dan *Selective Embedding Adaptation* (SEA).
7. Metode *Explainable AI* yang digunakan dibatasi pada LIME tanpa membandingkan dengan metode interpretabilitas lain.
8. Ruang lingkup perbandingan eksperimen mencakup skenario *full fine-tuning* dan skenario ARL-LoRA^{+} pada ketiga model; tidak mencakup metode deteksi hoaks berbasis pendekatan lain.
9. Evaluasi performa model dibatasi pada metrik F1-*weighted*, F1 per kelas (hoaks dan *real*), standar deviasi, *Coefficient of Variation* (CV), jumlah *trainable parameter*, waktu pelatihan, penggunaan memori GPU puncak (*peak VRAM*), *Expected Calibration Error* (ECE), dan analisis *confusion matrix*. Eksperimen multi-*seed* (5 *random seed*) diterapkan untuk mengukur stabilitas kinerja dan menghindari klaim yang bergantung pada satu inisialisasi acak.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Adapun tujuan spesifik dari penelitian ini diantaranya:

1. Mengevaluasi dan membandingkan kinerja XLM-RoBERTa, mBERT, dan DistilmBERT pada skenario *cross-lingual few-shot hoax detection* ke Bahasa Melayu menggunakan metrik F1-*weighted*, F1 per kelas, standar deviasi, dan *Coefficient of Variation* (CV) pada *test set* ketiga bahasa

(Bahasa Indonesia, Inggris, dan Melayu), serta menentukan konfigurasi terbaik antara *full fine-tuning* dan ARL-LoRA$^+$ untuk setiap arsitektur.

2. Menganalisis efisiensi komputasi ARL-LoRA$^+$ dibandingkan *full fine-tuning* melalui pengukuran jumlah *trainable parameter*, waktu pelatihan, dan penggunaan memori GPU puncak, serta mengidentifikasi kontribusi marjinal setiap komponen (ARL, BitFit, SEA) terhadap total peningkatan kinerja melalui *ablation study* sistematis pada XLM-RoBERTa.
3. Mengukur stabilitas kinerja ARL-LoRA$^+$ melalui eksperimen multi-*seed* (lima *random seed*) dan dua mode *sampling* Bahasa Melayu (mode *fix* dan *resample*), serta mengidentifikasi faktor arsitektural yang menentukan efektivitas ARL-LoRA$^+$ berdasarkan perbandingan hasil pada arsitektur 12-*layer* (XLM-RoBERTa dan mBERT) dan 6-*layer* (DistilmBERT).
4. Menerapkan metode *Explainable AI* berbasis LIME untuk mengidentifikasi kata-kata dengan bobot *feature importance* tertinggi yang mendorong prediksi ke kelas hoaks atau *real* pada masing-masing bahasa, serta membandingkan pola linguistik dan strategi pengambilan keputusan yang dikembangkan oleh ARL-LoRA$^+$ dan *full fine-tuning* lintas ketiga model.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan memberikan kontribusi positif bagi pengembangan ilmu pengetahuan, baik secara teoretis maupun praktis. Adapun kontribusi penelitian ini meliputi aspek teoretis dan praktis sebagai berikut:

1. Manfaat Penelitian Teoritis

   Kontribusi akademis dari penelitian ini berkaitan dengan sumbangannya dalam memperluas kajian mengenai deteksi berita palsu berbasis *multilingual transformer*, khususnya pada integrasi efisiensi parameter dan interpretabilitas model. Rincian manfaat teoritis tersebut adalah:

   a. Menghasilkan kerangka PEFT hybrid (ARL-LoRA$^+$) yang mengintegrasikan tiga mekanisme komplementer pada level berbeda

(*input representation*, *transformation*, dan *output decision*), sebagai kontribusi metodologis baru dalam literatur deteksi hoaks multilingual.

b. Memperluas pemahaman mengenai efektivitas *Adaptive-Rank LoRA* yang menerapkan rank berbeda per kelompok lapisan untuk menjaga keseimbangan antara preservasi representasi universal dan adaptasi *task-specific*.

c. Menyediakan bukti faktual mengenai kontribusi *Selective Embedding Adaptation* (SEA) dalam meningkatkan kualitas *cross-lingual transfer* ke bahasa serumpun melalui adaptasi embedding selektif pada level token bahasa target.

d. Menghadirkan integrasi evaluasi performa dan interpretabilitas melalui penerapan LIME, sehingga penelitian ini tidak hanya berfokus pada akurasi tetapi juga transparansi dan akuntabilitas model dalam mendeteksi hoaks lintas bahasa.

2. Manfaat Penelitian Praktis

Manfaat praktis penelitian ini berfokus pada penerapan hasil analisis untuk mendukung kebutuhan sistem moderasi konten digital dan pengambilan keputusan berbasis data. Melalui pendekatan deteksi hoaks multilingual yang efisien dan interpretable, penelitian ini diharapkan memberikan solusi aplikatif sebagai berikut:

a. Membantu institusi media, platform digital, maupun lembaga pemeriksa fakta dalam mendeteksi berita palsu lintas bahasa secara lebih akurat dan efisien menggunakan sampel data yang terbatas (*few-shot*).

b. Memfasilitasi pengembangan sistem moderasi konten otomatis yang mampu bekerja pada lingkungan dengan keterbatasan komputasi melalui pemanfaatan model *distilled* dan teknik ARL-LoRA^{+} yang hanya melatih kurang dari 2% parameter total model.

c. Menyediakan kerangka analitik berbasis AI yang tidak hanya memberikan prediksi klasifikasi, tetapi juga penjelasan terhadap

keputusan model, sehingga meningkatkan kepercayaan pengguna terhadap sistem deteksi hoaks.

d. Mendorong penerapan *AI-driven misinformation detection* pada konteks regional Asia Tenggara, khususnya dalam menghadapi dinamika distribusi berita lintas Bahasa Indonesia, Inggris, dan Melayu.

1.5 Sistematika Penulisan

Dibawah ini merupakan sistematika penulisan yang runtut pada setiap bab penelitian ini:

BAB I: PENDAHULUAN

Bab pertama berisi penjelasan mengenai latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan. Bab ini memberikan gambaran umum mengenai urgensi deteksi hoaks lintas bahasa dengan pendekatan *few-shot*, proposal metode ARL-LoRA^{+} sebagai pendekatan PEFT hybrid, serta kontribusi yang diharapkan.

BAB II: LANDASAN TEORI

Pada bab kedua, dibahas mengenai penelitian-penelitian sebelumnya yang relevan dengan topik deteksi berita palsu, *multilingual learning*, serta perkembangan model berbasis *transformer*. Teori-teori yang mendukung penelitian ini meliputi konsep dasar *Fake News Detection*, *Multilingual Transformer*, *Parameter-Efficient Fine-Tuning* termasuk *Adaptive-Rank LoRA*, BitFit, dan *Selective Embedding Adaptation*, serta *Explainable Artificial Intelligence* (XAI).

BAB III: METODOLOGI PENELITIAN

Bab ketiga menguraikan kerangka kerja KDD yang diterapkan dalam penelitian, mencakup lima tahap berurutan. Tahap *selection* menjelaskan sumber dan karakteristik *dataset* tiga bahasa yang digunakan. Tahap *pre-processing* mencakup beberapa langkah pembersihan teks dan *balanced sampling*. Tahap *transformation* menjelaskan pembagian data dengan skema *few-shot* untuk Bahasa Melayu,

identifikasi token SEA, dan tokenisasi spesifik per model. Tahap *modeling* mendeskripsikan konfigurasi *full fine-tuning* dan ARL-LoRA$^+$ beserta seluruh *hyperparameter*-nya. Tahap *evaluation* menjelaskan metrik yang digunakan, protokol eksperimen multi-*seed*, dan prosedur analisis LIME.

BAB IV: ANALISIS DAN HASIL PENELITIAN

Bab keempat menyajikan temuan dari seluruh tahap eksperimen secara terperinci. Bagian *selection* mendeskripsikan *dataset* yang digunakan beserta statistik distribusinya. Bagian *pre-processing* melaporkan hasil setiap langkah pembersihan data. Bagian *transformation* menjelaskan distribusi data setelah *splitting* dan hasil identifikasi token SEA. Bagian *modeling* menyajikan hasil pelatihan per model beserta *learning curve* dan *confusion matrix* per bahasa. Bagian *evaluation* memuat *gap analysis* kinerja lintas model, efisiensi komputasi, *ablation study* komponen ARL-LoRA$^+$, analisis sensitivitas *sampling*, analisis LIME per bahasa dan model, protokol *reproducibility*, serta diskusi komparatif dengan penelitian terdahulu.

BAB V: SIMPULAN DAN SARAN

Bab kelima berisi kesimpulan dari penelitian yang dilakukan untuk menjawab pertanyaan penelitian, serta saran-saran untuk penelitian selanjutnya terkait pengembangan model deteksi hoaks *multilingual* yang lebih adaptif, efisien, dan transparan.