Federated learning untuk Deteksi Hoaks pada Bahasa Daerah Bersumber Daya Rendah di Indonesia

Brilliani Jeshia Potalangi[1], Kristania Keren Bullu[2], Ade Yusupa[3], Ir. Yaulie Deo Y. Rindengan[4]
Program Studi Teknik Informatika, Universitas Sam Ratulangi, Indonesia
Corresponding Author: kristaniabullu026@student.unsrat.ac.id, brillianipotalangi026@student.unsrat.ac.id, ade@unsrat.ac.id

Abstrak

Penyebaran konten hoaks di berbagai platform media sosial menjadi ancaman serius bagi komunitas daerah di Indonesia, khususnya bagi mereka yang berkomunikasi menggunakan bahasa daerah bersumber daya rendah seperti Melayu Manado. Sistem deteksi hoaks yang ada masih terpusat pada Bahasa Indonesia standar, sehingga komunitas bahasa daerah sangat rentan terhadap disinformasi digital. Pendekatan pembelajaran mesin terpusat juga menimbulkan kekhawatiran privasi serius dalam konteks regulasi perlindungan data di Indonesia. Penelitian ini mengusulkan kerangka kerja berbasis *federated learning* untuk deteksi hoaks pada bahasa Melayu Manado yang mengintegrasikan Multilingual BERT (mBERT) sebagai arsitektur representasi teks, algoritma FedAvg, mekanisme DP-SGD untuk perlindungan privasi diferensial, dan framework Flower. Dataset terdiri dari 5.000 sampel berlabel yang dianotasi oleh penutur asli. Pelatihan *federated* dilaksanakan pada 3–5 node klien simulasi selama 20 putaran komunikasi. Hasil eksperimen menunjukkan bahwa model mBERT *federated* mencapai F1-score 87,6% pada distribusi data IID, hanya 1,8 poin di bawah baseline terpusat (89,4%), sekaligus mempertahankan jaminan privasi formal dengan anggaran privasi $\varepsilon \approx 8$.

Kata Kunci: Federated learning; Deteksi Hoaks; NLP Bahasa Daerah; Melayu Manado; Differential Privacy

Abstract

Hoax content spread across social media platforms poses a serious threat to regional communities in Indonesia, particularly those communicating in low-resource regional languages such as Manado Malay. Existing hoax detection systems remain focused on Standard Indonesian, leaving regional communities highly vulnerable to digital misinformation, while centralized machine learning approaches raise serious privacy concerns. This study proposes a federated learning–based framework for hoax detection in Manado Malay, integrating Multilingual BERT (mBERT), the FedAvg algorithm, DP-SGD for differential privacy, and the Flower framework. A dataset of 5,000 annotated samples was constructed and training was conducted across 3–5 simulated client nodes over 20 communication rounds. The federated mBERT model achieves an F1-score of 87.6% under IID data distribution, only 1.8 percentage points below the centralized baseline (89.4%), while maintaining formal privacy guarantees ($\varepsilon \approx 8$).

Keywords*: Federated learning; Hoax detection; Regional language NLP; Manado Malay; Differential privacy*

PENDAHULUAN

Indonesia merupakan salah satu negara dengan keanekaragaman bahasa terbesar di dunia, dengan lebih dari 700 bahasa daerah yang masih digunakan di berbagai wilayah kepulauan (Grimes,

2000). Meskipun demikian, sebagian besar penelitian dan perangkat pemrosesan bahasa alami (NLP) masih dikembangkan untuk Bahasa Indonesia standar atau bahasa besar dunia seperti Inggris dan Mandarin. Bahasa daerah seperti Melayu Manado, Jawa, Sunda, Minangkabau, dan Batak masih sangat kurang terwakili dalam riset linguistik komputasional, sehingga menciptakan kesenjangan besar dalam penanganan ancaman digital di komunitas tersebut (Wibowo, 2022).

Penyebaran konten hoaks melalui media sosial seperti WhatsApp, Facebook, dan TikTok telah menimbulkan gangguan sosial yang serius di Indonesia, khususnya di komunitas daerah yang literasi media kritisnya masih rendah (Kominfo, 2023). Sistem deteksi hoaks yang ada, yang umumnya dibangun di atas model *deep learning* terpusat, hanya mampu menangani Bahasa Indonesia dan tidak mampu menangkap pola linguistik, perilaku *code-switching*, serta nuansa semantik yang khas dalam komunikasi berbahasa daerah (Pratama et al., 2022). Selain itu, pendekatan terpusat juga mengharuskan data mentah pengguna dikirimkan ke server pusat, yang menimbulkan kekhawatiran terhadap privasi dan kedaulatan data dalam kerangka UU No. 27 Tahun 2022 tentang Perlindungan Data Pribadi (DPR-RI, 2022).

Penelitian-penelitian sebelumnya terkait deteksi hoaks di Indonesia umumnya menggunakan pendekatan *centralized learning* berbasis model seperti LSTM, CNN, atau fine-tuning BERT pada Bahasa Indonesia standar (Pratama & Suhartono, 2022). Meskipun penelitian terkini telah mengeksplorasi *federated learning* untuk deteksi berita palsu pada bahasa bersuber daya tinggi (Machová et al., 2024). (Wang et al., 2024) terdapat gap signifikan dalam literatur: (1) belum ada sistem deteksi hoaks berbasis *federated learning* yang secara khusus dirancang untuk bahasa daerah Indonesia bersumber daya rendah; (2) sebagian besar pendekatan *federated* yang ada belum mengintegrasikan jaminan privasi diferensial formal; dan (3) evaluasi pada kondisi data non-IID yang lebih mencerminkan realitas distribusi data terdesentralisasi masih sangat terbatas. Penelitian ini secara langsung menjawab ketiga gap tersebut dengan mengusulkan sistem berbasis *federated learning* yang dirancang khusus untuk konteks bahasa Melayu Manado, dilengkapi dengan perlindungan privasi melalui DP-SGD, dan dievaluasi secara komprehensif pada konfigurasi IID maupun non-IID.

Tujuan utama penelitian ini bersifat tiga arah: (1) mengembangkan sistem deteksi hoaks yang andal dan sadar bahasa untuk konteks bahasa daerah bersuber daya rendah; (2) mengimplementasikan arsitektur *federated learning* yang menjaga privasi sesuai regulasi perlindungan data Indonesia; dan (3) mengevaluasi secara empiris *trade-off* performa antara pendekatan pelatihan *federated* dan terpusat dalam konteks klasifikasi teks multibahasa bersuber daya rendah.

METODE

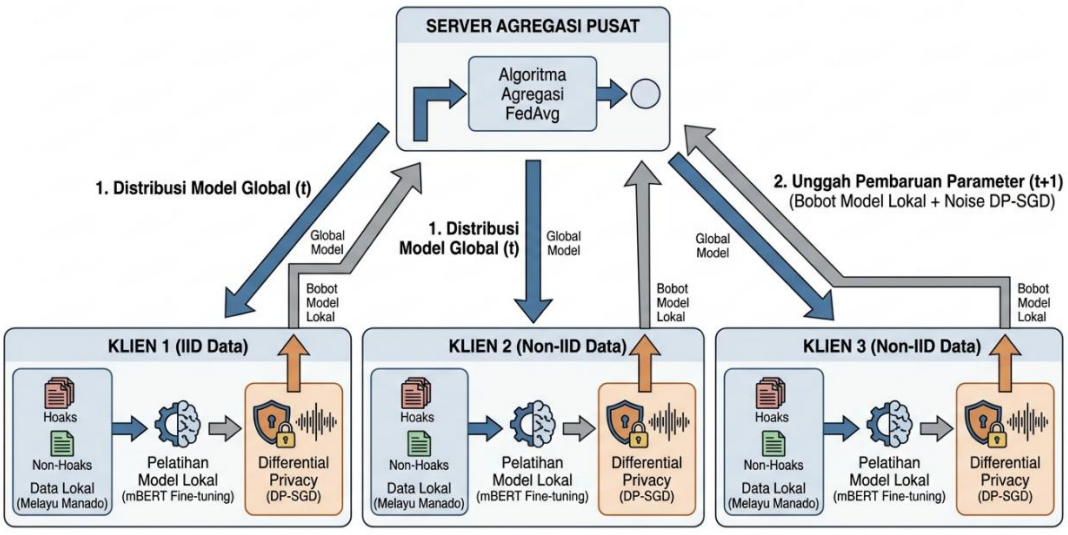
Penelitian ini menggunakan metodologi design science yang dikombinasikan dengan penelitian eksperimental machine learning. Alur kerja keseluruhan terdiri dari empat fase utama: (1) pengumpulan dan pra-pemrosesan dataset, (2) pemilihan model dasar dan pengaturan pelatihan lokal, (3) implementasi kerangka kerja *federated learning*, serta (4) evaluasi dan analisis performa.

Spesifikasi Perangkat Keras dan Lingkungan Eksperimen. Seluruh eksperimen dilaksanakan pada lingkungan komputasi dengan spesifikasi sebagai berikut: CPU Intel Core i7-12700H (14

core), RAM 32 GB DDR5, GPU NVIDIA GeForce RTX 3060 6 GB VRAM, dan penyimpanan SSD NVMe 512 GB. Sistem operasi yang digunakan adalah Ubuntu 22.04 LTS. Setiap node klien simulasi dijalankan sebagai proses Python independen menggunakan antarmuka virtual Flower, sehingga mensimulasikan komunikasi antar perangkat terdistribusi. Total waktu pelatihan untuk satu siklus penuh 20 putaran komunikasi pada 5 node klien adalah sekitar 4–6 jam bergantung pada konfigurasi DP-SGD. Model mBERT memiliki sekitar 179 juta parameter, sesuai dengan spesifikasi arsitektur multilingual-bert-base-uncased dari HuggingFace Transformers (Devlin et al., 2019).

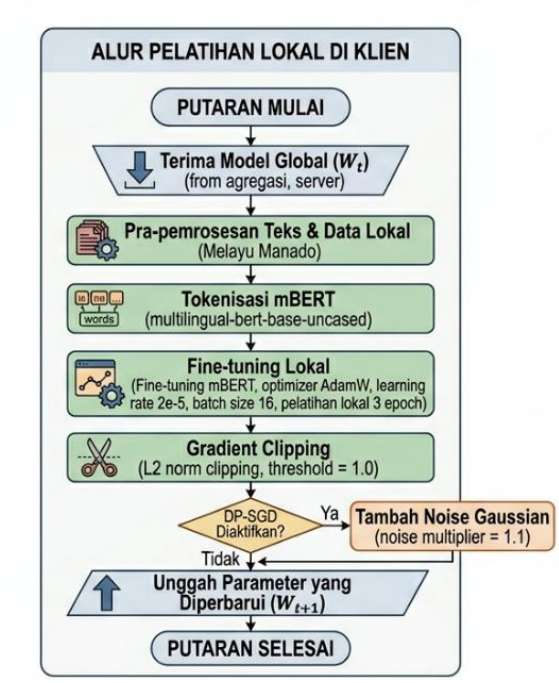
Pra-pemrosesan Teks. Sebelum memasuki tahap pelatihan, seluruh teks Melayu Manado menjalani serangkaian langkah pra-pemrosesan: (1) normalisasi Unicode dan konversi huruf kecil; (2) penghapusan karakter khusus, URL, dan emoji yang tidak relevan secara semantik; (3) segmentasi token menggunakan WordPiece tokenizer bawaan mBERT dengan panjang maksimal 128 token per sampel; (4) padding dan truncation untuk memastikan dimensi input yang seragam; serta (5) pelabelan ulang melalui verifikasi konsistensi anotasi. Mengingat karakteristik bahasa Melayu Manado yang mengandung code-switching dengan Bahasa Indonesia dan unsur bahasa lokal, tidak dilakukan penghapusan stopword secara agresif untuk mempertahankan konteks semantik yang relevan.

Skenario Simulasi Node Klien. Simulasi federated learning diimplementasikan menggunakan 3 hingga 5 node klien virtual yang masing-masing merepresentasikan komunitas daerah yang berbeda. Node klien 1 (1.500 sampel) mensimulasikan komunitas pengguna WhatsApp perkotaan di Manado. Node klien 2 (1.200 sampel) mensimulasikan komunitas media sosial (Facebook/TikTok) di wilayah Minahasa Utara. Node klien 3 (1.100 sampel) mensimulasikan komunitas yang data utamanya bersumber dari repositori Kominfo. Node klien 4 dan 5 (masing-masing 600 sampel) diaktifkan pada eksperimen 5-klien untuk mengevaluasi skalabilitas agregasi. Pada konfigurasi IID, setiap klien menerima partisi acak yang seimbang. Pada konfigurasi non-IID, distribusi data di setiap klien mencerminkan kekhasan sumber dan subdialek masing-masing komunitas, sehingga reprodusibel oleh peneliti lain dengan menggunakan random seed 42 yang ditetapkan secara konsisten.



Gambar 1. Arsitektur Sistem

Dalam arsitektur sistem yang diusulkan, beberapa node klien yang mewakili komunitas daerah berbeda melatih model mBERT lokal secara independen menggunakan dataset masing-masing. Setiap klien kemudian mengunggah hanya parameter model (bobot) bukan data mentah ke server agregasi pusat. Server menerapkan algoritma FedAvg untuk menghitung rata-rata berbobot dari parameter yang diterima, menghasilkan model global yang didistribusikan kembali ke seluruh klien untuk putaran pelatihan berikutnya. Proses ini berulang untuk sejumlah putaran komunikasi yang telah ditentukan. Injeksi noise Gaussian untuk differential privacy (DP-SGD) diterapkan di sisi klien sebelum pengunggahan parameter.



Gambar 2. Alut Pelatihan

Penelitian sebelumnya telah menunjukkan kelayakan *federated learning* untuk deteksi berita palsu menggunakan arsitektur deep learning seperti LSTM dan CNN (Wibowo, 2022), yang mengkonfirmasi bahwa pelatihan terdesentralisasi tidak secara signifikan menurunkan performa klasifikasi dibandingkan pendekatan terpusat. Untuk pengaturan multibahasa bersuber daya rendah, model berbasis transformer seperti mBERT dan XLM-RoBERTa menunjukkan kemampuan transfer lintas bahasa yang kuat (Wang et al., 2024). Selain itu, integrasi differential privacy dengan klasifikasi teks *federated* berbasis transformer telah terbukti mencapai trade-off privasi-utilitas yang menguntungkan (Chowdhury et al., 2021).

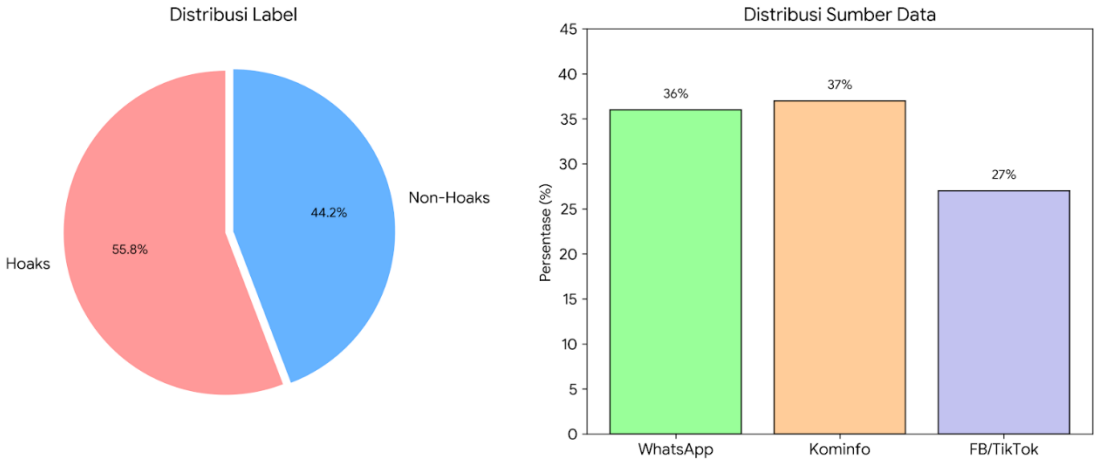
HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Dataset

Dataset yang dikompilasi dalam penelitian ini terdiri dari 5.000 sampel teks berlabel yang bersumber dari tiga saluran berbeda: platform media sosial termasuk Facebook dan TikTok (1.350 sampel), arsip percakapan grup WhatsApp (1.800 sampel), dan basis data publik Kominfo Hoaks Busters (1.850 sampel). Setiap sampel dianotasi secara independen oleh minimal tiga penutur asli bahasa Melayu Manado dan diberikan label klasifikasi biner — hoaks atau non-hoaks — menggunakan protokol adjudikasi majority voting. Dataset yang dihasilkan mengandung 2.790 sampel berlabel hoaks (55,8%) dan 2.210 sampel berlabel non-hoaks (44,2%), mencerminkan ketidakseimbangan kelas yang sedikit namun dapat dikelola.

Tabel 1. Komposisi Dataset berdasarkan Sumber dan Distribusi Label

Sumber	Label: Hoaks	Label: Non-Hoaks	Total Sampel
Arsip WhatsApp	980	820	1.800
Postingan Facebook/TikTok	710	640	1.350
Kominfo Hoaks Busters	1.100	750	1.850
Total	**2.790**	**2.210**	**5.000**



Gambar 3. Distribusi Dataset *(Pie/Bar Chart)*

Keterangan Visual:

- Distribusi Label: Menunjukkan proporsi kelas Hoaks sebesar 55,8% (2.790 sampel) dan Non-Hoaks sebesar 44,2% (2.210 sampel).
- Distribusi Sumber Data: Visualisasi kontribusi dari tiga saluran utama yaitu WhatsApp (36%), Kominfo Hoaks Busters (37%), dan platform media sosial Facebook/TikTok (27%).

Dataset kemudian dipartisi menggunakan rasio 80:10:10 untuk subset pelatihan, validasi, dan pengujian. Eksperimen dilakukan dalam dua konfigurasi distribusi data: IID (independent and identically distributed), di mana setiap klien menerima sampel acak yang seimbang secara statistik; dan non-IID, di mana data didistribusikan berdasarkan kategori sumber dan variasi subdialek daerah, sehingga lebih mencerminkan heterogenitas nyata dari deployment *federated* di dunia nyata.

3.2 Pengaturan Eksperimental

Seluruh eksperimen diimplementasikan dalam Python 3.10 menggunakan framework PyTorch untuk pelatihan model dan library Flower (flwr) 1.5.0 untuk koordinasi *federated learning*. Model pre-trained multilingual-bert-base-uncased yang tersedia melalui HuggingFace Transformers digunakan sebagai arsitektur dasar. Pelatihan lokal di setiap node klien dilakukan selama 3 epoch per putaran komunikasi menggunakan optimizer AdamW dengan learning rate 2e-5 dan batch size 16. Pelatihan *federated* dilaksanakan selama 20 putaran komunikasi global. Untuk eksperimen differential privacy, DP-SGD dikonfigurasi dengan noise multiplier sebesar 1,1 dan threshold clipping norma L2 sebesar 1,0, menghasilkan anggaran privasi formal epsilon (ε) $\approx$ 8 di bawah kerangka Rényi Differential Privacy (RDP) (Mironov, 2017)

3.3 Hasil Performa Klasifikasi

Tabel 2 menyajikan ringkasan performa klasifikasi secara komprehensif di semua konfigurasi eksperimental, dievaluasi pada set uji menggunakan akurasi, presisi, recall, dan F1-score sebagai metrik utama. Baseline mBERT terpusat, yang dilatih pada dataset gabungan lengkap menggunakan fine-tuning supervised konvensional, mencapai performa tertinggi dengan F1-score 89,4%, yang menjadi referensi batas atas untuk konfigurasi *federated*.

Tabel 2. Perbandingan Performa Klasifikasi di Semua Konfigurasi Eksperimental

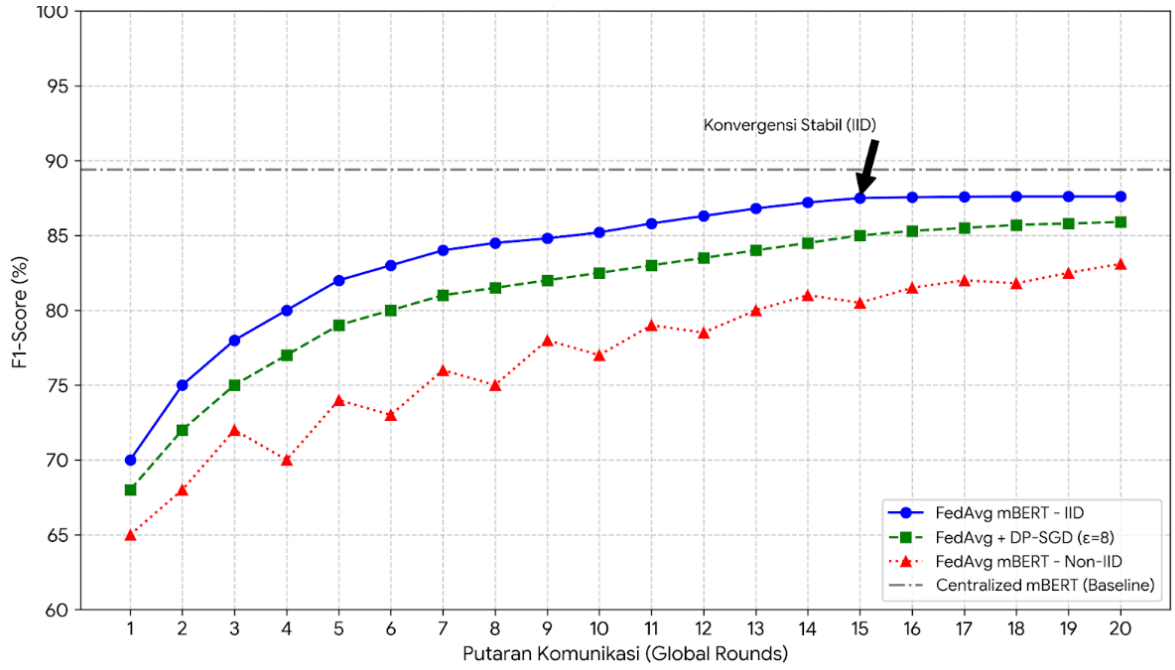
Konfigurasi Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
Centralized mBERT (Baseline)	89,4	88,7	90,1	89,4
FedAvg mBERT - IID (Putaran 10)	85,2	84,9	85,6	85,2
FedAvg mBERT - IID (Putaran 20)	87,6	87,0	88,3	87,6
FedAvg mBERT - Non-IID (Putaran 20)	83,1	82,4	83,9	83,1
FedAvg + DP-SGD (ε=8, Putaran 20)	85,9	85,2	86,7	85,9

Pada distribusi data IID, model mBERT *federated* yang dilatih selama 20 putaran komunikasi mencapai F1-score 87,6%, dengan selisih performa hanya 1,8 poin persentase dibandingkan baseline terpusat. Perbedaan yang sangat kecil ini mengkonfirmasi bahwa pelatihan *federated* tidak menyebabkan penurunan efektivitas klasifikasi yang substansial ketika data terdistribusi secara merata. Dalam kondisi non-IID yang lebih akurat mencerminkan distribusi heterogen dari data terdesentralisasi dunia nyata performa menurun ke F1-score 83,1%, konsisten dengan temuan penelitian *federated learning* sebelumnya pada distribusi data heterogen (Li, 2022).

Ketika batasan *differential privacy* diintegrasikan melalui DP-SGD ($\varepsilon \approx 8$) ke dalam konfigurasi pelatihan *federated* IID, model mencapai F1-score 85,9% setelah 20 putaran komunikasi. Hasil ini menunjukkan bahwa jaminan differential privacy yang bermakna dapat diperoleh sekaligus mempertahankan performa klasifikasi pada level yang sangat praktis untuk deployment aplikasi deteksi hoaks bahasa daerah. Trade-off privasi-utilitas yang diamati konsisten dengan temuan (Chowdhury et al., 2021)

3.4 Analisis Konvergensi

Pemantauan performa model global di berbagai putaran komunikasi menunjukkan bahwa konfigurasi *federated* IID mencapai konvergensi stabil pada sekitar putaran ke-15, setelah itu peningkatan marjinal berkurang secara substansial. Sebaliknya, konfigurasi non-IID menunjukkan perilaku konvergensi yang lebih lambat dan tidak stabil, dengan osilasi F1-score validasi yang bertahan hingga putaran pelatihan akhir. Observasi ini memperkuat pemahaman bahwa heterogenitas data merupakan tantangan mendasar dalam optimasi *federated* dan memotivasi eksplorasi lebih lanjut tentang strategi personalized *federated learning* (Li, 2022).



Gambar 4. Grafik Konvergensi *F1-Score* per Putaran Komunikasi

Analisis Visual:

- Stabilitas IID: Garis biru (IID) menunjukkan kenaikan yang konsisten dan mulai mencapai titik jenuh (stabil) pada sekitar putaran ke-15 dengan skor akhir 87,6%.
- Osilasi Non-IID: Garis merah (Non-IID) mengalami fluktuasi yang lebih tinggi dan konvergensi yang lebih lambat, berakhir di angka 83,1%, mencerminkan tantangan heterogenitas data.
- Efek DP-SGD: Garis hijau (IID + DP-SGD) mengikuti pola IID namun dengan sedikit keterlambatan dan skor akhir 85,9% akibat adanya injeksi noise untuk perlindungan privasi.
- Garis Baseline: Garis putus-putus abu-abu di atas menunjukkan batas performa model terpusat (Centralized) sebagai referensi pembanding sebesar 89,4%.

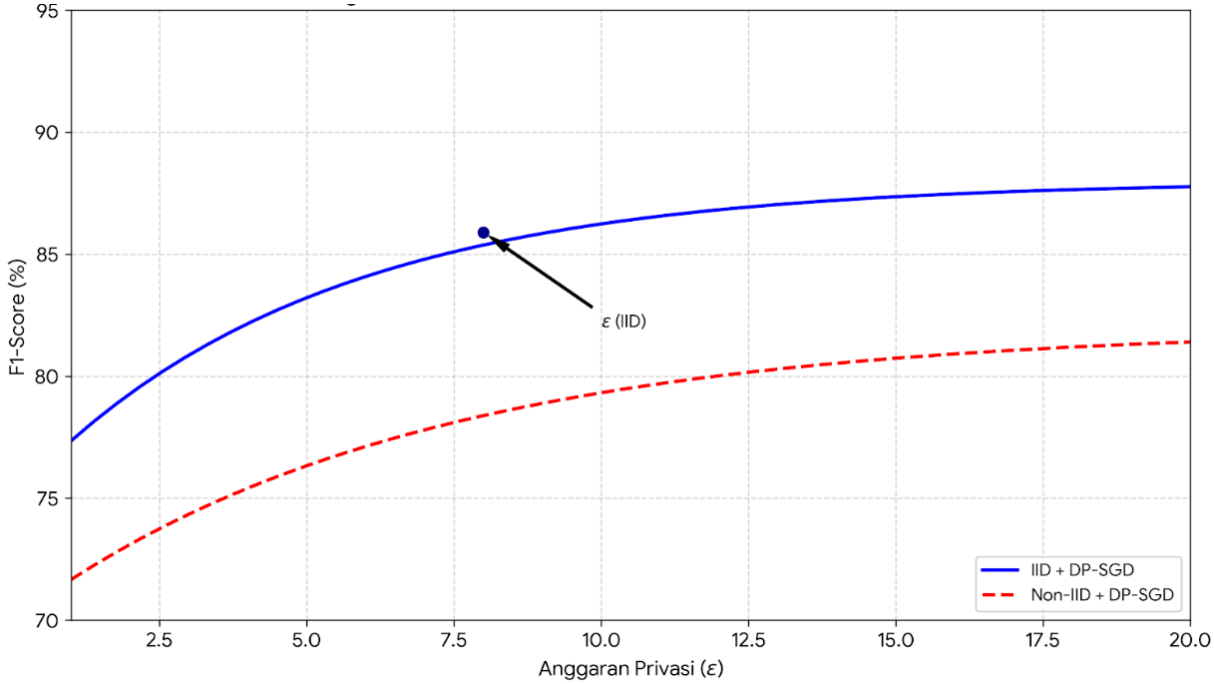
3.5 Perbandingan dengan Penelitian Terkait

Performa yang dicapai oleh model mBERT *federated* dalam penelitian ini secara umum konsisten dan dalam beberapa aspek lebih baik dibandingkan hasil yang dilaporkan dalam penelitian serupa sebelumnya. (Machová et al., 2024) melaporkan F1-score deteksi berita palsu *federated* dalam kisaran 81–86% menggunakan arsitektur berbasis LSTM, sementara pendekatan berbasis transformer dalam penelitian ini mencapai 87,6% dalam kondisi IID, mencerminkan keunggulan performa dari representasi multibahasa pre-trained (Wang et al., 2024). Sejauh pengetahuan penulis, penelitian ini merupakan penerapan pertama *federated learning* untuk deteksi hoaks khususnya dalam konteks Melayu Manado dan bahasa daerah Indonesia yang sebanding, sehingga memberikan kontribusi baru yang belum pernah diteliti sebelumnya dalam lanskap riset NLP bahasa bersuber daya rendah di Indonesia.

3.6 Analisis Privasi

Analisis privasi formal dilakukan dalam kerangka Rényi Differential Privacy (RDP) (Mironov, 2017), dengan biaya privasi per putaran dilacak menggunakan metodologi moment accountant. Selama 20 putaran pelatihan dengan hyperparameter DP-SGD yang ditentukan (noise multiplier

= 1,1, L2 clip norm = 1,0), anggaran privasi kumulatif dihitung sebagai ε ≈ 8 pada nilai delta (δ) sebesar 10^{-5}, yang merupakan jaminan privasi yang secara praktis masuk akal dan dapat diinterpretasikan secara hukum dalam konteks UU Perlindungan Data Pribadi Indonesia (DPR-RI, 2022). Secara kritis, data mentah pengguna dari sumber media sosial dan arsip WhatsApp pribadi tidak pernah dikirimkan di luar node klien masing-masing dalam tahapan pelatihan mana pun.



Gambar 5. Kurva *Trade-off* Privasi-Utilitas (ε vs *F1-Score)*

Analisis Visual:

- Hubungan Positif: Sesuai dengan temuan penelitian, terdapat hubungan positif antara ϵ dan F1-score; saat anggaran privasi meningkat (batasan privasi dilonggarkan), performa klasifikasi meningkat secara monoton.
- Titik Optimal: Pada pengaturan eksperimental $\epsilon \approx 8$, konfigurasi IID+DP-SGD berhasil mencapai F1-score 85,9%, yang dianggap sebagai *trade-off* yang praktis dan menguntungkan antara perlindungan data dan akurasi.
- Dampak Heterogenitas: Kurva Non-IID secara konsisten berada di bawah kurva IID pada semua nilai ϵ, yang menunjukkan bahwa perbedaan distribusi data di tingkat klien tetap menjadi faktor penentu utama kualitas model *federated*.

Baik konfigurasi IID maupun non-IID menunjukkan hubungan positif yang konsisten antara ε dan F1-score: seiring meningkatnya anggaran privasi yang diizinkan (yaitu, ketika batasan privasi dilonggarkan), performa klasifikasi meningkat secara monoton. Konfigurasi IID mencapai F1-score yang jauh lebih tinggi dibandingkan non-IID di semua nilai ε, mencerminkan dampak persisten dari heterogenitas data terhadap kualitas model *federated*. Pada pengaturan eksperimental ε = 8, konfigurasi IID+DP-SGD mencapai F1-score 85,9%, yang merupakan trade-off menguntungkan antara perlindungan privasi dan utilitas klasifikasi.

3.7 Implikasi Praktis dan Tantangan Deployment

Meskipun hasil eksperimen menunjukkan performa yang kompetitif, implementasi sistem ini di lingkungan nyata menghadapi sejumlah tantangan yang perlu dipertimbangkan secara serius. Pertama, terkait keterbatasan bandwidth: setiap putaran komunikasi federated membutuhkan transmisi parameter model mBERT (~1,4 GB per klien), yang menjadi hambatan signifikan di komunitas daerah dengan infrastruktur jaringan terbatas. Optimasi seperti kompresi gradient, kuantisasi model, atau penggunaan arsitektur yang lebih ringan seperti DistilBERT perlu dieksplorasi lebih lanjut untuk konteks ini.

Kedua, terkait keamanan komunikasi antar node: meskipun DP-SGD melindungi privasi data individual, sistem federated tetap rentan terhadap serangan seperti model poisoning dan Byzantine attacks dari klien yang tidak terpercaya (Li et al., 2022). Implementasi di dunia nyata perlu mempertimbangkan mekanisme deteksi anomali dan verifikasi integritas parameter sebelum agregasi. Penggunaan saluran komunikasi terenkripsi (TLS/SSL) antara klien dan server agregasi juga wajib diterapkan.

Ketiga, kesiapan infrastruktur komunitas lokal merupakan tantangan yang tidak kalah penting. Sistem ini memerlukan perangkat dengan kapasitas komputasi memadai di sisi klien untuk menjalankan fine-tuning mBERT secara lokal. Bagi komunitas Melayu Manado yang sebagian besar menggunakan perangkat smartphone kelas menengah, adaptasi arsitektur ke model yang lebih ringan atau pendekatan split learning perlu dikaji. Selain itu, edukasi dan keterlibatan komunitas dalam proses anotasi data serta pemahaman tentang manfaat privasi dari sistem federated menjadi faktor kritis keberhasilan implementasi jangka panjang.

SIMPULAN

Penelitian ini telah mengusulkan dan memvalidasi secara empiris kerangka kerja *federated learning* untuk deteksi hoaks dalam bahasa Melayu Manado. Sistem yang dikembangkan mengintegrasikan mBERT, FedAvg, dan DP-SGD dalam satu pipeline yang menjamin privasi data pengguna sesuai regulasi UU No. 27 Tahun 2022.

Hasil eksperimen menunjukkan tiga temuan utama: (1) model mBERT *federated* mencapai F1-score 87,6% pada kondisi IID, hanya 1,8 poin di bawah baseline terpusat; (2) integrasi DP-SGD ($\varepsilon \approx 8$) mempertahankan F1-score 85,9%, mengkonfirmasi bahwa perlindungan privasi yang bermakna dapat diberikan tanpa mengorbankan utilitas klasifikasi secara substansial; dan (3) pada kondisi non-IID yang lebih realistis, F1-score menurun ke 83,1%, menyoroti tantangan heterogenitas data sebagai arah penelitian lanjutan yang kritis.

Kontribusi utama penelitian ini meliputi: (1) penerapan pertama *federated learning* untuk deteksi hoaks pada bahasa daerah Indonesia bersumber daya rendah; (2) demonstrasi empiris keseimbangan antara performa kompetitif dan jaminan *differential privacy* formal; serta (3) penetapan kerangka metodologis yang dapat direplikasi untuk komunitas bahasa daerah lainnya di Indonesia. Penelitian lanjutan disarankan untuk mengeksplorasi arsitektur yang lebih ringan, strategi personalized *federated learning* untuk mengatasi heterogenitas data, serta implementasi pada perangkat yang benar-benar terdistribusi secara geografis.

UCAPAN TERIMA KASIH

Para penulis menyampaikan penghargaan yang tulus kepada pembimbing akademik penelitian ini, Ade Yusupa (ade@unsrat.ac.id) dan Ir. Yaulie Deo Y. Rindengan (rindengan@unsrat.ac.id), atas bimbingan

ahli, masukan konstruktif, dan mentorship yang berkelanjutan selama proses penelitian dan penulisan naskah. Para penulis juga berterima kasih kepada para sukarelawan penutur asli bahasa Melayu Manado yang berkontribusi dalam upaya anotasi dataset, serta kepada Program Studi Teknik Informatika Universitas Sam Ratulangi atas sumber daya komputasional dan lingkungan akademis yang mendukung penelitian ini. Penelitian ini dilaksanakan sebagai bagian dari persyaratan mata kuliah Riset Informatika dalam Program Studi Teknik Informatika Universitas Sam Ratulangi.

DAFTAR PUSTAKA

Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Parcollet, T., de Guênet de Montjoye, P. P. B., & Lane, N. D. (2020). *Flower: A friendly federated learning research framework.* arXiv preprint. https://arxiv.org/abs/2007.14390

Chowdhury, S., Dubey, S., & Singh, M. (2021). *Privacy enabled financial text classification using differential privacy and federated learning.* arXiv preprint. https://arxiv.org/abs/2110.01643

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding.* In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT) (pp. 4171–4186). Association for Computational Linguistics.

Dewan Perwakilan Rakyat Republik Indonesia. (2022). *Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi.* Sekretariat Negara Republik Indonesia.

Grimes, B. F. (2000). *Ethnologue: Languages of the world* (14th ed.). SIL International.

Kementerian Komunikasi dan Informatika Republik Indonesia. (2023). *Klarifikasi berita hoaks.* https://www.kominfo.go.id/content/all/laporan_isu_hoaks

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Smola, A., & Smith, V. (2022). *Federated learning on non-IID data silos: An experimental study.* In Proceedings of the 38th IEEE International Conference on Data Engineering (ICDE) (pp. 965–978). IEEE. https://doi.org/10.1109/ICDE53745.2022.00077

Machová, K., Mach, M., & Balara, V. (2024). *Federated learning in the detection of fake news using deep learning as a basic method.* Sensors, 24(11), Article 3590. https://doi.org/10.3390/s24113590

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). *Communication-efficient learning of deep networks from decentralized data.* In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) (pp. 1273–1282). PMLR.

Mironov, I. (2017). *Rényi differential privacy of the Gaussian mechanism.* In Proceedings of the 30th IEEE Computer Security Foundations Symposium (CSF) (pp. 263–275). IEEE. https://doi.org/10.1109/CSF.2017.11

Plank, B., Søgaard, A., & Goldberg, Y. (2016). *Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss.* In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 412–418). Association for Computational Linguistics.

Pratama, A., & Suhartono, D. (2022). *Hoax detection in Indonesian social media using deep learning.* Journal of ICT Research and Applications,

16(1), 1–18. https://doi.org/10.5614/itbj.ict.res.appl.2022.16.1.1

Rust, P., Pfeiffer, J.-S., Vulić, I., Ruder, S., & Gurevych, I. (2021). *How good is your tokenizer? On the monolingual performance of multilingual language models.* In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 4011–4025). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021

Wang, X., Zhang, W., & Rajtmajer, S. (2024). *Monolingual and multilingual misinformation detection for low-resource languages: A comprehensive survey*. arXiv preprint. https://arxiv.org/abs/2410.18390

Wibowo, R., Kurniawan, A., & Suhartono, D. (2022). *Low-resource NLP for Indonesian regional languages: Challenges and opportunities.* In Proceedings of the ACL Workshop on NLP for Low-Resource Languages (pp. 45–54). Association for Computational Linguistics.