

IMPLEMENTASI FEDERATED LEARNING UNTUK PREDIKSI PRESTASI MAHASISWA TANPA BERBAGI DATA PRIBADI

Yogi Eriawan¹, Stevanus Immanuel Buulolo²

¹ Program Studi Informatika Program Sarjana Universitas Audi Indonesia

² Mahasiswa Prodi informatika Program Sarjana

chrisdayanti1103@gmail.com

Abstrak

Perkembangan Artificial Intelligence (AI) dan Machine Learning telah mendorong pemanfaatan teknologi dalam bidang pendidikan, salah satunya untuk memprediksi prestasi mahasiswa. Namun, metode pembelajaran terpusat (*Centralized Learning*) mengharuskan pengumpulan data akademik dari berbagai institusi ke satu server sehingga berpotensi menimbulkan risiko terhadap keamanan dan privasi data mahasiswa. Penelitian ini bertujuan untuk mengimplementasikan Federated Learning sebagai solusi dalam membangun model prediksi prestasi mahasiswa tanpa melakukan pertukaran data pribadi antar institusi, sehingga keamanan dan kerahasiaan data tetap terjaga.

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen. Data yang digunakan berupa data akademik mahasiswa yang meliputi Indeks Prestasi Kumulatif (IPK), kehadiran, nilai tugas, nilai Ujian Tengah Semester (UTS), nilai Ujian Akhir Semester (UAS), serta aktivitas mahasiswa pada Learning Management System (LMS). Setiap institusi melakukan pelatihan model secara lokal menggunakan data yang dimiliki, kemudian parameter model dikirimkan ke server pusat untuk dilakukan proses agregasi menggunakan algoritma Federated Averaging (FedAvg). Model global yang dihasilkan kemudian dievaluasi menggunakan metrik Accuracy, Precision, Recall, dan F1-Score.

Hasil penelitian menunjukkan bahwa implementasi Federated Learning mampu menghasilkan model prediksi prestasi mahasiswa dengan performa yang baik tanpa mengharuskan perpindahan data akademik antar institusi. Pendekatan ini tidak hanya mendukung kolaborasi dalam pengembangan model prediksi, tetapi juga mampu menjaga privasi data mahasiswa karena data tetap tersimpan pada masing-masing institusi. Dengan demikian, Federated Learning dapat menjadi solusi yang efektif dalam pengembangan sistem prediksi prestasi mahasiswa yang aman, kolaboratif, dan mendukung pengambilan keputusan akademik berbasis data.

Kata Kunci: *Federated Learning*, prediksi prestasi mahasiswa, *Machine Learning*, *Educational Data Mining*, *Federated Averaging*, privasi data..

Implementation of Federated Learning for Student Academic Performance Prediction Without Sharing Personal Data

Abstract

The rapid development of Artificial Intelligence (AI) and Machine Learning has significantly influenced the education sector, particularly in predicting students' academic performance. However, conventional Centralized Learning requires academic data from multiple institutions to be collected on a central server, which raises concerns regarding data privacy and security. This study aims to implement Federated Learning as a privacy-preserving approach for predicting students' academic performance without requiring the sharing of personal data among participating institutions.

This research employed a quantitative approach using an experimental research method. The dataset consisted of students' academic records, including Grade Point Average (GPA), attendance, assignment scores, midterm examination scores, final examination scores, and Learning Management System (LMS) activity. Each participating institution trained the prediction model locally using its own dataset, while only the model parameters were transmitted to a central server for aggregation using the Federated Averaging (FedAvg)

algorithm. The performance of the global model was evaluated using Accuracy, Precision, Recall, and F1-Score metrics.

The results demonstrate that the implementation of Federated Learning is capable of producing an effective student academic performance prediction model while preserving the privacy of academic data. Since students' data remain within each institution and only model parameters are exchanged, the proposed approach reduces privacy risks without compromising predictive performance. Therefore, Federated Learning offers a secure and collaborative solution for developing academic performance prediction systems and supports data-driven decision-making in higher education institutions.

Keywords: Federated Learning, Student Academic Performance Prediction, Machine Learning, Educational Data Mining, Federated Averaging, Data Privacy.

PENDAHULUAN

Perkembangan teknologi Artificial Intelligence (AI) telah memberikan dampak signifikan terhadap transformasi digital di sektor pendidikan. Salah satu implementasi AI adalah pemanfaatan Machine Learning untuk menganalisis data akademik dan memprediksi prestasi mahasiswa. Prediksi prestasi mahasiswa sangat penting sebagai dasar penyusunan strategi pembelajaran, pemberian pendampingan akademik, serta pengambilan keputusan yang lebih efektif oleh perguruan tinggi.

Sebagian besar penelitian sebelumnya menggunakan pendekatan *Centralized Learning*, yaitu menggabungkan seluruh data akademik ke dalam satu server pusat. Pendekatan tersebut mampu menghasilkan model prediksi yang baik, tetapi memiliki kelemahan berupa meningkatnya risiko kebocoran data dan pelanggaran privasi mahasiswa. Kondisi ini menjadi tantangan karena data akademik merupakan informasi yang bersifat sensitif dan harus dilindungi.

Federated Learning hadir sebagai pendekatan yang memungkinkan beberapa institusi melatih model secara kolaboratif tanpa memindahkan data mentah. Pada metode ini, data tetap tersimpan di masing-masing institusi, sedangkan server hanya menerima parameter model yang telah diperbarui. Dengan demikian, keamanan dan kerahasiaan data dapat dipertahankan.

Penelitian ini bertujuan mengimplementasikan Federated Learning menggunakan algoritma **Federated Averaging (FedAvg)** untuk membangun model prediksi prestasi mahasiswa yang tetap menjaga privasi data akademik.

METODE PENELITIAN

2.1 Jenis Penelitian

Penelitian ini menggunakan metode penelitian kuantitatif dengan pendekatan eksperimen. Pendekatan ini bertujuan untuk mengimplementasikan Federated Learning dalam membangun model prediksi prestasi mahasiswa tanpa melakukan pertukaran data pribadi antar institusi. Selanjutnya, performa model dievaluasi menggunakan beberapa metrik, yaitu Accuracy, Precision, Recall, dan F1-Score.

2.2 Lokasi dan Objek Penelitian

Penelitian dilaksanakan pada beberapa perguruan tinggi yang berperan sebagai client dalam arsitektur Federated Learning. Masing-masing institusi menyimpan data akademik mahasiswa secara lokal dan hanya mengirimkan parameter model ke server pusat..

2.3 Tahapan Penelitian

Adapun tahapan penelitian yang dilakukan adalah sebagai berikut.

1. Studi Literatur, yaitu mengumpulkan dan mempelajari berbagai referensi berupa buku, jurnal ilmiah, prosiding, dan artikel yang berkaitan dengan Artificial Intelligence, Machine Learning, Federated Learning, Educational Data Mining, serta prediksi prestasi mahasiswa.
2. Pengumpulan Data Akademik Mahasiswa, yaitu mengumpulkan data yang digunakan dalam penelitian, seperti Indeks Prestasi Kumulatif (IPK), kehadiran, nilai tugas, nilai Ujian Tengah Semester (UTS), nilai Ujian Akhir Semester (UAS), dan aktivitas mahasiswa pada Learning Management System (LMS).
3. Pra-pemrosesan Data (*Data Preprocessing*), yaitu membersihkan data dari nilai yang hilang (*missing value*), menghapus data duplikat, melakukan normalisasi, serta mengubah data kategorikal menjadi data numerik agar siap digunakan dalam proses pelatihan model.
4. Pembagian Data pada Masing-masing Client, yaitu mendistribusikan data ke setiap institusi atau *client*. Pada tahap ini, data tetap tersimpan pada masing-masing *client* dan tidak dipindahkan ke server pusat.
5. Pelatihan Model Lokal (*Local Training*), yaitu setiap *client* melakukan proses pelatihan model menggunakan data lokal yang dimiliki sehingga menghasilkan parameter model masing-masing.
6. Agregasi Model Menggunakan Algoritma Federated Averaging (FedAvg), yaitu server pusat menggabungkan parameter model yang diperoleh dari setiap *client* untuk membentuk model global tanpa menerima data mentah dari masing-masing institusi.
7. Pembentukan Model Global, yaitu menghasilkan model global berdasarkan proses agregasi yang kemudian didistribusikan kembali kepada seluruh *client* untuk digunakan pada proses pelatihan berikutnya hingga model mencapai kondisi optimal.
8. Evaluasi Performa Model, yaitu mengukur kinerja model menggunakan metrik evaluasi Accuracy, Precision, Recall, dan F1-Score untuk mengetahui tingkat ketepatan model dalam memprediksi prestasi mahasiswa.
9. Analisis Hasil Penelitian, yaitu menganalisis hasil evaluasi model untuk mengetahui efektivitas implementasi Federated Learning dalam memprediksi prestasi mahasiswa sekaligus menjaga keamanan dan privasi data akademik. Hasil analisis kemudian digunakan sebagai dasar dalam penarikan kesimpulan dan penyusunan rekomendasi penelitian.

2.4 Arsitektur Federated Learning

Alur implementasi terdiri atas:

1. Server Menginisialisasi Model Global

Tahap pertama dalam implementasi Federated Learning adalah server melakukan inisialisasi model global yang akan digunakan sebagai model awal dalam proses pelatihan

2. Model Global Dikirim ke Seluruh Client

Setelah model global berhasil diinisialisasi, server mendistribusikan model tersebut kepada seluruh *client* yang berpartisipasi dalam proses Federated Learning. Model yang dikirim berupa struktur model beserta parameter awal yang akan digunakan sebagai dasar pelatihan pada masing-masing *client*

3. Masing-masing Client Melakukan Pelatihan Menggunakan Data Lokal

Data yang digunakan meliputi Indeks Prestasi Kumulatif (IPK), kehadiran, nilai tugas, nilai Ujian Tengah Semester (UTS), nilai Ujian Akhir Semester (UAS), dan aktivitas mahasiswa pada *Learning Management System* (LMS)

4. Parameter Model Lokal Dikirim Kembali ke Server

Setelah proses pelatihan lokal selesai, setiap *client* mengirimkan parameter model yang telah diperbarui kepada server pusat. Parameter tersebut berupa bobot (*weights*) dan bias hasil pembelajaran model terhadap data lokal.

5. Server Melakukan Agregasi Menggunakan Algoritma Federated Averaging (FedAvg)

Algoritma ini menggabungkan parameter model lokal menjadi satu model global baru dengan mempertimbangkan jumlah data yang dimiliki oleh masing-masing *client*.

6. Model Global Terbaru Dikirim Kembali ke Seluruh Client

Setelah proses agregasi selesai, server menghasilkan model global terbaru yang memiliki parameter hasil kombinasi dari seluruh model lokal.

7. Proses Diulang Hingga Jumlah Round Pelatihan Terpenuhi atau Model Mencapai Konvergensi

Tahapan distribusi model, pelatihan lokal, pengiriman parameter, dan agregasi dilakukan secara berulang dalam beberapa putaran (*communication rounds*). Setiap putaran bertujuan untuk memperbarui model global agar memiliki kemampuan prediksi yang semakin baik.

2.5 Metode Federated Averaging (FedAvg)

Metode agregasi yang digunakan adalah **Federated Averaging (FedAvg)** yang diperkenalkan oleh McMahan et al. (2017).

Persamaan FedAvg :

$$W_{global} = \sum_{k=1}^K \frac{n_k}{n} W_k$$

Keterangan :

- W_{global} = bobot model global
- W_k = bobot model client ke-k
- n_k = jumlah data pada client ke-k
- n = total seluruh data
- K = jumlah client

Client	Jumlah Data	Bobot Model
Client A	100	0,75
Client B	200	0,80
Client C	300	0,90

Jumlah seluruh data :

$$n = 100 + 200 + 300 = 600$$

Perhitungan model global :

$$\begin{aligned}
 W_{global} &= \frac{100}{600}(0,75) + \frac{200}{600}(0,80) + \frac{300}{600}(0,90) \\
 &= 0,1667(0,75) + 0,3333(0,80) + 0,5(0,90) \\
 &= 0,125 + 0,267 + 0,450 \\
 W_{global} &= 0,842
 \end{aligned}$$

Hasil agregasi menunjukkan bahwa bobot model global yang diperoleh adalah **0,842**

2.5 Evaluasi Model

Kinerja model dievaluasi menggunakan **Confusion Matrix**

Tabel Confusion Matrix

	Prediksi Positif	Prediksi Negatif
Aktual Positif	TP	FN
Aktual Negatif	FP	TN

Keterangan :

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

1. Accuracy

Accuracy menunjukkan tingkat ketepatan model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

2. Precision

Precision menunjukkan ketepatan prediksi positif

$$Precision = \frac{TP}{TP + FP}$$

3. Recall

Recall menunjukkan kemampuan model menemukan data positif

$$Recall = \frac{TP}{TP + FN}$$

4. F1-Score

F1-Score merupakan rata-rata harmonis antara Precision dan Recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

2.6 Perhitungan Evaluasi

Misalkan diperoleh hasil pengujian:

- TP = 90
- TN = 80
- FP = 10
- FN = 2

a. Accuracy

$$\begin{aligned} &= \frac{90 + 80}{90 + 80 + 10 + 20} \\ &= \frac{170}{200} = 0,85 = 85\% \end{aligned}$$

b. Precision

$$= \frac{90}{90 + 10} = 0,90 = 90\%$$

c. Recall

$$= \frac{90}{90 + 20} = 0,818 = 81,8\%$$

d. F1-Score

$$= 2 \times \frac{0,90 \times 0,818}{0,90 + 0,818} = 0,857 = 85,7\%$$

2.8 Teknik Analisis Data

1. Analisis data dilakukan dengan membandingkan hasil prediksi model Federated Learning terhadap data aktual mahasiswa. Performa model dianalisis berdasarkan nilai Accuracy, Precision, Recall, dan F1-Score. Selain itu, penelitian juga

mengevaluasi kemampuan Federated Learning dalam menjaga privasi data karena data akademik tetap berada pada masing-masing institusi dan tidak dikirimkan ke server pusat. Hasil evaluasi digunakan untuk menilai efektivitas metode Federated Learning dalam membangun model prediksi prestasi mahasiswa yang akurat, aman, dan mendukung kolaborasi antar perguruan tinggi.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Federated Learning berhasil diterapkan untuk membangun model prediksi prestasi mahasiswa tanpa berbagi data pribadi antar institusi, sehingga privasi data tetap terjaga.
2. Algoritma Federated Averaging (FedAvg) mampu menggabungkan parameter model dari setiap *client* menjadi model global tanpa memerlukan pengiriman data mentah.
3. Model yang dihasilkan menunjukkan performa yang baik berdasarkan metrik Accuracy, Precision, Recall, dan F1-Score dalam memprediksi prestasi mahasiswa.
4. Penerapan Federated Learning mendukung pengembangan sistem analitik pendidikan yang aman, kolaboratif, dan berorientasi pada perlindungan data akademik.
5. Penelitian ini menunjukkan bahwa **Federated Learning** berpotensi menjadi solusi untuk mendukung pengambilan keputusan akademik berbasis data dengan tetap menjaga keamanan dan privasi informasi mahasiswa..

DAFTAR PUSTAKA

Stallings, W. (2020). *Effective Cybersecurity: A Guide to Using Best Practices and Standards*. Addison-Wesley.

Whitman, M. E., & Mattord, H. J. (2022). *Principles of Information Security* (7th ed.). Cengage.

NIST. (2024). *Cybersecurity Framework (CSF) 2.0*. National Institute of Standards and Technology.

Stallings, W. (2020). *Effective Cybersecurity: A Guide to Using Best Practices and Standards*. Addison-Wesley.

Whitman, M. E., & Mattord, H. J. (2022). *Principles of Information Security* (7th ed.). Cengage LearningKompas. (2022, 15 Mei). Gudang mi instan kadaluarsa di Semarang digerebek BPOM

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>

Creswell, J. W., & Creswell, J. D. (2022). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). *Multivariate data analysis* (9th ed.). Cengage Learning.

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2021). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, 4(1), Article 20. <https://doi.org/10.1186/s42400-021-00088-7>

Müller, A. C., & Guido, S. (2022). *Introduction to machine learning with Python: A guide for data scientists* (2nd ed.). O'Reilly Media.

Stallings, W. (2020). *Effective cybersecurity: A guide to using best practices and standards*. Addison-Wesley.

Whitman, M. E., & Mattord, H. J. (2022). *Principles of information security* (7th ed.). Cengage Learning.