Analisis Performa *Federated Learning* Untuk Klasifikasi Gambar Yang Sensitif Terhadap Privasi Data

1st I Made Krisna Bayu
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
krisnnabayu@students.telkomuniversity.ac.id

2nd Anggunmeka Luhur Prasasti
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
anggunmeka@telkomuniversity.ac.id

3rd Dicky Maryopi
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
maryopi@telkomuniversity.ac.id

Abstrak

**Pada Saat ini *machine learning* sudah digunakan di segala bidang yang berhubungan dengan aplikasi maupun perangkat keras yang bertujuan untuk membantu pengguna untuk melakukan tugas dengan campur tangan manusia yang minimal. Data merupakan kunci penting dari machine learning untuk diolah agar mesin mampu memberikan perintah sesuai dengan keadaan dan kondisi yang sedang dihadapi oleh mesin. Data tersebut bisa berupa data umum maupun data pribadi yang rawan disebarluaskan. Maka perlindungan data pribadi milik klien sangat penting dan harus dilindungi dengan cara mengaplikasikan *Federated Learning*. Pada penelitian ini mendapatkan hasil dari evaluasi terhadap *Federated Learning.* Hasil evaluasi pada *Federated Learning* didapatkan nilai rata-rata akurasi dari 20 kali pelatihan model pada *datacenter* adalah sebesar 70% dan rata-rata nilai loss sebesar 1.04. Penelitian ini juga mendapatkan hasil perbandingan *Federated Learning* dan *machine learning* tradisional dengan hasil menunjukkan bahwa akurasi *Federated Learning* lebih rendah dibanding dengan reguler *machine learning*. Serta nilai *loss,* dan *error* yang lebih tinggi pada *Federated Learning* dibandingkan dengan *Machine learning* tradisional. Dilakukan juga evaluasi akurasi pada *Federated Learning* dengan mengubah jumlah klien dan jumlah pelatihan model pada klien dan menunjukkan bahwa jumlah pelatihan model pada perangkat klien mempengaruhi kualitas akurasi model pada
pelatihan di *datacenter Federated Learning*.**

Kata Kunci: *Machine learning*, kecerdasan buatan, *Federated learning,* data pribadi, Performa, klasifikasi gambar.evaluasi.

Abstract

At this time machine learning has been used in all fields related to applications and hardware which aims to help users to perform tasks with minimal human intervention. Data is an important key for machine learning to be processed so that machines can give orders according to the circumstances and conditions being faced by the machine. The data can be in the form of public data or personal data that is prone to be disseminated. So, the protection of the client's personal data is very important and must be protected by applying Federated Learning. In this study, the results obtained from the evaluation of Federated Learning. The results of the evaluation on Federated Learning showed that the average accuracy value of 20 times the model training at the datacenter was 70% and the average loss value was This study also obtained the results of a comparison of Federated Learning and traditional machine learning with the results showing that the accuracy of Federated Learning is lower than regular machine learning. As well as higher loss and error values in Federated Learning compared to traditional machine learning. An evaluation of the accuracy of Federated Learning is also carried out by changing the number of clients and the number of model training on clients and showing that the number of model training on client devices affects the quality of model accuracy in training in the Federated Learning datacenter.

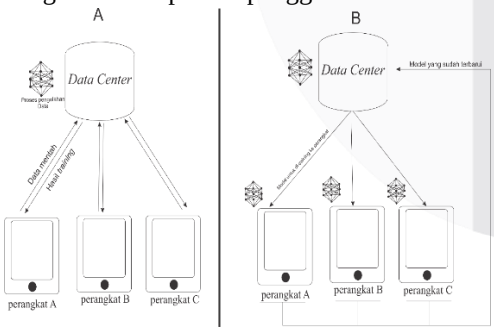
Keywords: Machine learning, Artificial intelligence, Federated learning, personal data, Performance, Imageclassification. Evaluation.

I. PENDAHULUAN

Pada saat ini, aplikasi yang kita temukan sehari-hari sudah mengimplementasikan *machine learning* untuk mengolah data yang akan digunakan baik oleh pengguna maupun developer. Data ini merupakan kunci utama dari *machine learning* untuk melatih model yang digunakan pada *data center*. Maka dari itu tidak aneh lagi jika ada transaksi data antara pengguna dan developer. Dengan metode *federated learning* yang memiliki kemampuan untuk menghindari transaksi data antara pengguna dengan *data center* [1]. Di penelitian ini penulis akan melakukan evaluasi performa dari *machine learning* yang menggunakan dan tidak menggunakan metode *federated learning* pada aplikasi klasifikasi gambar. Di akhir penelitian akan muncul variabel yang menjadi acuan untuk mengukur performa dari kedua metode yang akan di evaluasi. Pada penelitian ini juga dilakukan pengukuran akurasi dengan jumlah klien dan jumlah pelatihan pada setiap klien yang berbeda.

II. KAJIAN TEORI

Federated Learning (FL) adalah jenis distribusi data ML yang di mana kumpulan data mampu di-*training* dengan cara desentralisasi atau tanpa harus mengirim data mentah dari perangkat ke *data center*. Tujuan desentralisasi ini adalah untuk menjaga privasi pemilik data [1]. ML tradisional masih menggunakan sentralisasi pelatihan model menyebabkan kerahasiaan dari pemilik data tidak terjaga dengan baik sesuai dengan hukum privasi pengguna.



Gambar 2. 1 (A) *Machine learning* tradisional, (B) *federated learning*

Sesuai dengan ilustrasi dari gambar 2.1, FL bekerja dengan cara *data center* memberikan model awal setiap klien. Setiap klien akan melakukan pelatihan data pada perangkat masing-masing dan akan mengirim model parameter untuk memperbaharui model global. Dengan hanya mengirim parameter model dari gabungan beberapa perangkat maka data lokal pada perangkat tetap terisolasi tanpa berpindah ke *data center* [2].

a. TensorFlow

TensorFlow adalah suatu *framework* dari Google yang khusus dibuat untuk *Machine Learning. TensorFlow* memiliki fitur *open-source* yang bernama TensorFlow Federated (TFF) yang berfungsi sebagai *framework* pendukung ML dan komputasi lain yang berhubungan dengan desentralisasi data. Dengan TFF, melatih suatu data untuk menjadi model akan lebih mudah apabila sedang mengaplikasikan FL [3]. TFF memungkinkan developer untuk menyimulasi algoritma FL di model dan data mereka.

b. Keras

Keras adalah API *deep learning* yang ditulis dengan bahasa Python dan berjalan pada *framework* TensorFlow. Keras merupakan *interface* dengan *highly-productive* untuk memecahkan masalah pada *machine learning*, dengan berfokus pada *deep learning* modern. Pengembangan Keras memiliki tujuan pada kecepatan eksperimen. Keras sangat simpel karena pengembangan Keras bertujuan untuk memfokuskan peneliti pada masalah yang diberikan [4].

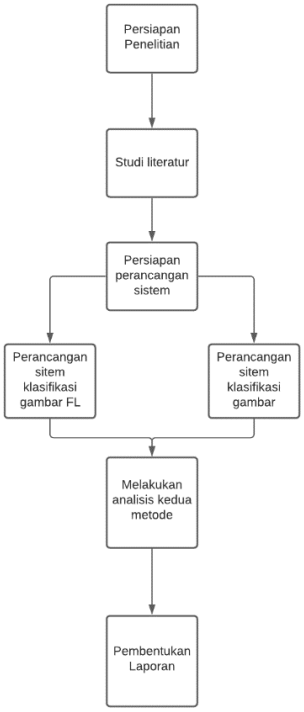
c. MNist

Mnist merupakan kumpulan data untuk tulisan tangan angka yang biasanya digunakan pada *machine learning* untuk dijadikan *dataset*. MNist memiliki 60.000 sampel dan 10.000 sampel tes [5]. Pada eksperimen ini data MNist sudah di *pre-processing* sebelumnya dengan Leaf agar data dibagi-bagi setiap klien dan mempermudah melakukan simulasi pada *federated learning*. Berikut sampel model yang dibentuk pada klien.

Gambar 2. 2 Sampel model klien.

III. METODE

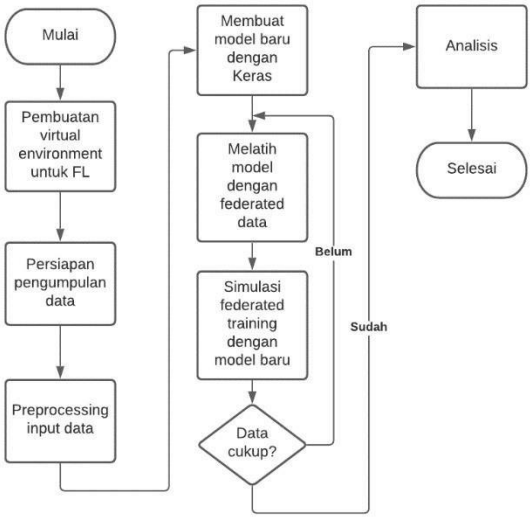
Berikut prosedur yang dilakukan peneliti untuk melakukan evaluasi terhadap *Federated Learning*



Gambar 3. 1 *Prosedur Penelitian.*

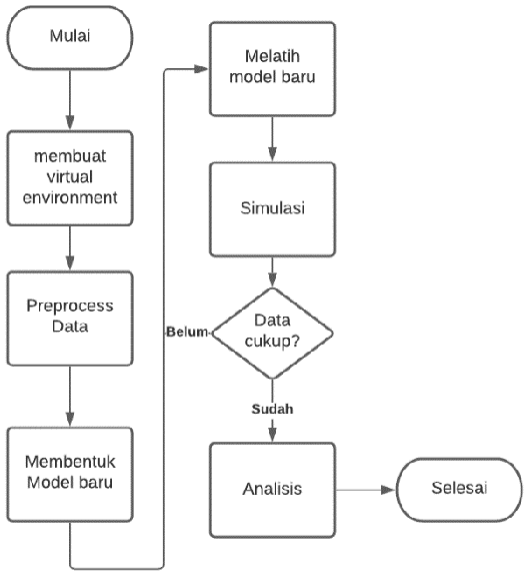
Pada penelitian ini dibentuk 2 sistem yaitu sistem *Federated learning* dan sistem *machine learning* tradisional yang nanti akan di komparasi hasil evaluasi antara kedua sistem tersebut,

Gambar 3.2 adalah pembentukan sistem pada *Federated Learning*



Gambar 3. 2 Desain sistem *federated learning*

Federated learning membutuhkan *federated data* set yang dikumpulkan dari hasil dari pelatihan pada perangkat klien yang nanti akan dilakukan pelatihan data lagi agar menjadi *global model* pada *datacenter*. Tahap ini tidak terjadi pada *machine learning* tradisional karena pelatihan hanya dilakukan sekali pada *datacenter*. Seperti digambarkan pada gambar 3.3.



Gambar 3. 3 Desain sistem machine learning.

IV. HASIL DAN PEMBAHASAN

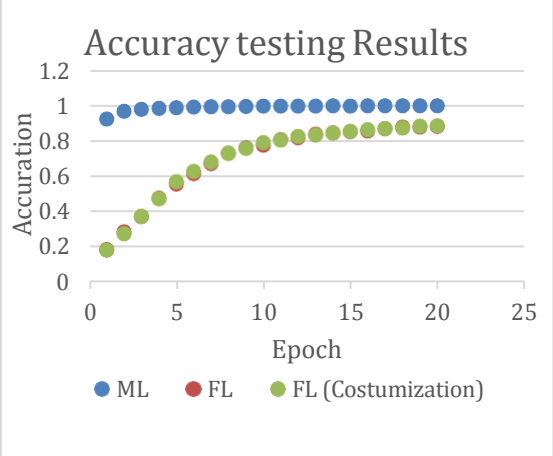
Hasil dari eksperimen ini adalah untuk mengukur performa dari *federated learning*. Aplikasi yang menjadi media mengukur adalah klasifikasi gambar menggunakan tulisan tangan angka. Adapun yang diukur adalah akurasi, dan data *loss* pada *federated learning*. Dari akurasi bisa diukur *mean absolute error* dan *mean squared error* yang nanti akan menghasilkan jumlah *error* pada *federated learning*. Pengukuran yang sama akan dilakukan pada *machine learning* tradisional yang nantinya menjadi nilai komparasi dari *federated learning*.

1. Akurasi

Akurasi dan *loss* akan dilakukan dengan pengukuran *multi-class cross-entropy*. Metode ini sangat membantu terutama pada klasifikasi yang memiliki kelas yang banyak seperti tulisan tangan angka. Mengukur akurasi dengan cara seperti ini

$$Akurasi = \frac{Prediksi\ benar}{Jumlah\ data} \quad (4.1) \quad [6]$$

Pada simulasi ini, peneliti menggunakan 20 kali epoch. Epoch adalah proses pelatihan model yang berhasil. Sebelum pelatihan pada *datacenter*, *federated learning* melakukan 20 kali epoch juga untuk dilakukan di setiap perangkat klien.



Gambar 4. 1 Diagram acak akurasi

Seperti diperlihatkan pada Gambar 4.21 Federated learning memiliki akurasi yang rendah pada percobaan training pertama dibanding dengan machine learning tradisional. Machine learning tradisional memiliki nilai akurasi yang tinggi pada awal percobaan training model menunjukkan bahwa dengan sekali training saja sudah mampu mendapatkan hasil yang baik.

Pada gambar 4.1, bisa dapat disimpulkan bahwa Federated learning harus melakukan pelatihan model pada data center berkali-kali untuk mendapatkan hasil yang baik. Menurut percobaan pada penelitian ini dibutuhkan juga sistem komputasi yang baik di setiap kliennya agar memberikan model yang baik ke data center federated learning. Pada dunia nyata pengaplikasian federated learning biasanya dilakukan pada perangkat klien berada pada mode charging atau idle agar tidak mengganggu aktivitas klien pada saat menggunakan perangkatnya.

Tabel 4. 1 Hasil rata-rata akurasi 20 epoch.

ML	FL	FL(Customization)
99%	70%	70%

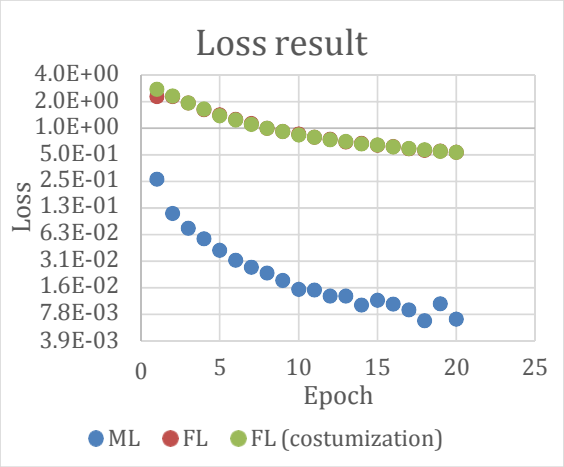
2. *Sparse Cross-entropy Loss*

Pada proses pelatihan data terjadi *Loss* atau kehilangan bobot model. Pada simulasi ini, *loss* bisa didapatkan dari mengaplikasikan metode penghitungan yang bernama *Sparse Categorical Crossentropy*. Menggunakan *sparse categorical crossentropy* karena pada simulasi ini dilakukan *multi-class* klasifikasi. Berikut adalah rumus menghitung *loss* dengan *sparse categorical crossentropy*.

$$J(w) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (4.2)$$

Dimana
w = model parameter
y_i= Label benar
$\hat{y}$= label prediksi [7]

Gambar 4.2 adalah hasil dari perhitungan *loss* 20 kali epoch.



Gambar 4. 2 hasil loss 20 epoch.

Pada diagram Gambar 4.2 menunjukkan bahwa setiap *training* menghasilkan *loss* data yang menurun dari *training* sebelumnya. *Loss* yang turun pada setiap pelatihannyanya diartikan sebagai mesin berjalan dengan baik. Pada *Federated learning* terjadi *loss* data yang sangat besar pada *training* pertama, namun *loss* berkurang setiap dilakukan *training* model. Maka bisa ditarik kesimpulan bahwa *Federated learning* membutuhkan *training* model pada *federated data* berkali-kali agar model yang dimiliki menjadi optimal.

Tabel 4. 2 Hasil *loss* rata-rata dari 20 epoch.

ML	FL	FL(customization)
0.0382	1.0543	1.0778

3. MSE dan MAE pada mesin yang dibentuk

Mean Absolute Error dan *Mean Squared Error* diambil dari data yang didapat pada Akurasi dari model yang sudah di-training. Angka MSE dan MAE apabila semakin besar maka error yang dimiliki pada mesin tersebut besar. Angkat MSE dan MAE yang paling baik adalah angka yang sangat kecil sehingga diketahui bahwa error yang terdapat pada mesin tersebut kecil. Menghitung MSE dan MAE menggunakan data akurasi yang dijabarkan pada Akurasi Training Model. Cara menghitung MSE-nya yaitu.

$$MAE = \frac{1}{N}\sum_{i=1}^{n} |y - \hat{y}| \quad (4.3)$$

N = Jumlah Data

y = Hasil prediksi

$\hat{y}$ = Hasil prediksi benar

Hasil dari perhitungan MAE adalah

Tabel 4. 3 MAE.

ML	FL	FL(Costumization)
0.01208	**0.303**	**0.302**

Hasil dari MAE tidak begitu relevan dalam kehidupan sehari-hari karena kita melakukan evaluasi terhadap model yang sama. Dengan kata lain, MAE akan menghasilkan hasil yang bagus jika bertemu data pada *dataset* dan hasil yang tidak bagus ketika bertemu data yang tidak ada di *dataset.*

Mean Squared Error memiliki rumus yang mirip seperti MAE. Berikut adalah rumus untuk menghitung MSE [8].

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{n} (hasil\ prediksi - prediksi\ benar)^2 \quad (4.4)$$

Tabel 4. 4 MSE.

ML	FL	FL (Customization)
0.00044	**0.135372**	**0.135037**

Pada tabel 4.3 dan tabel 4.4 ditunjukan bahwa *Error* pada FL tetap lebih besar dibandingkan dengan *machine learning* tradisional. *Error* ini akan berdampak pada hasil klasifikasi dari mesin yang dibuat. Penyebab dari tingginya *Error* pada FL adalah model yang didapat pada data *center* merupakan model hasil dari *training* pada setiap klien.

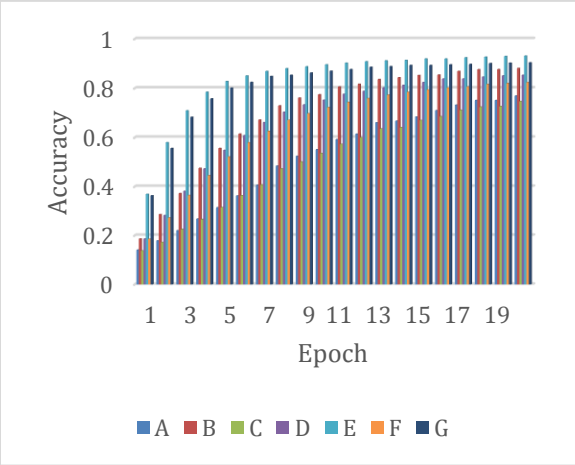
4. Mengubah pengulangan pelatihan pada FL

Peneliti melakukan analisis lainnya pada *Federated learning*. Analisis ini disimulasikan dengan jumlah klien dan jumlah *training* setiap klien yang berbeda. Agar adil, jumlah *batch* ditetapkan sama yaitu 20. Pada analisis ini dilakukan simulasi dengan data sebagai berikut.

Tabel 4. 5 Sampel Klien.

Sampel	Client	Client's Epoch	Data Center's Epoch
A	10	10	20
B	10	20	20
C	20	10	20
D	20	20	20
E	20	50	20
F	50	20	20
G	50	50	20

Dengan data seperti tabel 4.5 maka didapatkan akurasi seperti gambar 4.3



Gambar 4. 3 Hasil akurasi dari sampel.

Dengan rata-rata akurasi sebagai berikut

Tabel 4. 6 Rata-rata akurasi dari setiap sampel

A	B	C	D	E	F	G
52%	70%	51%	68%	84%	65%	82%

Dari percobaan ini dapat dilihat bahwa jumlah akurasi yang didapat berbeda untuk setiap percobaan yang dilakukan. Pada grafik Gambar 4.3 menunjukkan bahwa besar jumlah *training* pada klien mempengaruhi kualitas akurasi yang didapat pada *data center*. Pada percobaan antara C yang memiliki jumlah *training* model pada klien sebesar 10 kali dan E yang sebesar 50 kali terdapat perbedaan besar akurasi sebesar 33%.

Perbedaan jumlah klien yang berpartisipasi untuk menyumbang data tidak begitu berpengaruh besar terhadap jumlah akurasi model pada *data center*. Hal ini dibuktikan dengan perbandingan pada E dan G yang hanya memiliki perbedaan akurasi sebesar 2%. Dilihat dari nilainya perbedaan ini didapatkan dari semakin besar jumlah kliennya akan berpengaruh mengurangi akurasi model walau hanya sedikit.

Kualitas akurasi yang ditunjukkan pada grafik Gambar 4.3 menunjukkan bahwa jumlah *training model* pada setiap klien sangat mempengaruhi akurasi pada saat model di *training* kembali di *data center*. Maka, fenomena ini membuktikan bahwa *federated learning* membutuhkan klien yang melakukan *training model* pada perangkatnya masing-masing agar menghasilkan global model yang baik untuk diberikan ke *data center*.

V. KESIMPULAN

Berdasarkan penelitian dan pengujian yang telah dilaksanakan pada Tugas Akhir ini, maka dapat disimpulkan bahwa Federated Learning memiliki hasil akurasi yang rendah pada tahap pertama pelatihan model yaitu sebesar 18%. Pada saat 20 kali pelatihan model, FL memiliki rata-rata akurasi sebesar 70%. Pada loss data saat melakukan pelatihan model, FL juga memiliki hasil loss yang sangat besar yaitu 2.8 dan memliki rata-rata loss dari 20 pelatihan model sebesar 1.04.

Perbedaan performa antara *Federated Learning* dengan reguler *machine learning* sangat signifikan. Dilihat dari akurasi dengan pengulangan pelatihan data 20 kali. FL memiliki akurasi lebih rendah dibanding *machine learning* dengan perbandingan sebesar 29%. Pada *Loss*, FL memiliki 1.0161 lebih besar dibanding dengan *Machine learning*. Begitu juga nilai error model dengan metode MAE dan MSE, pada FL lebih besar dibanding dengan *regular machine learning*.

Dampak jumlah klien pada *Federated Learning* tidak terlalu berpengaruh terhadap akurasi dari pelatihan model di *data center*. Namun, jumlah pelatihan model pada setiap klien berpengaruh terhadap hasil akurasi dari pelatihan model di *data center*. Terdapat

REFERENSI

[1] Q. Yang, Federated Learning, China: Morgan & Claypool, 2019.

[2] J. H. Yoo, "Federated Learning: Issues in Medical Application," p. 3, 2021.

[3] N. Kourtellis, "FLaaS: Federated Learning as a Service," 2020.

[4] "About Keras," keras, [Online]. Available: https://keras.io/about/. [Accessed 29 September 2021].

[5] Y. LeCun, "The MNIST DATABASE," [Online]. Available: yann.lecun/exdb/mnist. [Accessed 3 2 2021].

[6] H. Kumar, "Loss vs Accuracy," 7 Desember 2018. [Online]. Available: kharshit.github.io/vkif/2018/12/07/loss-vs-accuracy. [Accessed 1 Februari 2021].

[7] K. E. Koech, "Cross-Entropy Loss Function," Toward Data Science, 3 Oktober 2020. [Online]. Available: https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e. [Accessed 28 Januari 2020].

[8] SmritiS, "What is Mean Squared Error, Mean Absolute Error, Root Mean Squared Error and R Squared?," Study Tonight, 10 agustus 2021. [Online]. Available: https://www.studytonight.com/post/what-is-mean-squared-error-mean-absolute-error-root-mean-squared-error-and-r-squared. [Accessed 2 Oktober 2021].

[9] K. Bonawitz, "TOWARDS FEDERATED LEARNING AT SCALE: SYSTEM DESIGN," p. 1, 2019.