A Client Selection Approach Based on Dynamic Clustering and Reputation in Federated Learning

Gabriel A. Fonseca[1], Arthur A. Brito[1], Paulo Vitor B. Ramos[2], Bruno H. Dias[2], Saulo M. Villela[1], Hugo Morais[3], Lucas Pereira[4]

[1]Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora
Juiz de Fora – MG – Brasil

[2]Departamento de Energia Elétrica – Universidade Federal de Juiz de Fora
Juiz de Fora – MG – Brasil

[3]INESC-ID – Universidade de Lisboa
Lisboa – Portugal

[4]LARSyS – Polo Científico e Tecnológico da Madeira
Funchal – Portugal

{andrioli.gabriel,arthuraugusto.brito,paulo.barbosa}@estudante.ufjf.br,

{bruno.dias,saulo.moraes}@ufjf.br,

{hugo.morais,lucas.pereira}@tecnico.ulisboa.pt

Abstract. *Electrical demand forecasting faces challenges related to data confidentiality and the heterogeneity of consumption profiles. This work proposes a clustered federated learning strategy that combines a cluster topology revised at runtime through merging and creation with a reputation function that weights the quality of local models alongside exploration bonuses for both clients and clusters. The results demonstrate that the cluster models achieved a mean squared error approximately 0.146 orders of magnitude lower than that of a single global model trained with FedAvg, while also better tracking the consumption curves of the participants within their respective clusters.*

Resumo. *A previsão de demanda elétrica enfrenta desafios de confidencialidade e de heterogeneidade dos perfis de consumo. Este trabalho propõe uma estratégia de aprendizado federado com agrupamento (clustered federated learning) que combina uma topologia de clusters revisada em tempo de execução por fusão e criação com uma função de reputação que pondera a qualidade dos modelos locais e bônus de exploração de cliente e de cluster. Os resultados demonstram que os modelos-cluster alcançaram erro quadrático médio cerca de 0,146 ordens de grandeza menor que o de um modelo global único treinado com FedAvg, além de acompanhar melhor a curvas dos participantes de seu cluster.*

1. Introdução

A crescente demanda global por energia, aliada à necessidade de descarbonização, impulsionou a transição para sistemas de energia distribuídos e renováveis, cada vez mais informatizados. Ao ganho de controlabilidade correspondeu um aumento significativo da

complexidade das operações em sistemas elétricos de potência [Thwe et al. 2025]. Nesse cenário, marcado pelo alto volume de dados, a previsão da demanda de eletricidade torna-se um fator crítico para garantir a estabilidade, a alocação eficiente de reservas de contingência e a mitigação de riscos operacionais [Casolaro et al. 2023].

A previsão de demanda está presente em diversos domínios do setor elétrico, como em sistemas com alta penetração de fontes intermitentes, prever a geração otimiza o despacho de fontes de resposta rápida, mitigando o risco de distúrbios sistêmicos [Azzalini et al. 2025]. Adicionalmente, a previsão é também determinante para estratégias de resposta à demanda [Navidi et al. 2023], como a integração de veículos elétricos [He 2024] e otimização de sistemas de armazenamento [Humiston et al. 2026]. No entanto, o comportamento histórico isolado não basta para predições confiáveis, sendo necessária a inclusão de múltiplas variáveis correlatas, processo que acarreta o aumento da dimensionalidade do problema e, portanto, do custo computacional. Consequentemente, abordagens tradicionais de previsão tornaram-se menos eficientes para fornecer resultados em tempo hábil, como exige a operação moderna [Ramos et al. 2023].

Modelos baseados em aprendizado profundo (*Deep Learning* – DL) consolidaram-se como estado da arte por sua capacidade de capturar correlações não-lineares complexas [Dakheel and Çevik 2025]. Entretanto, esses modelos enfrentam dois desafios críticos: a escassez de dados no cenário de *cold start*, em que novos ativos não possuem histórico suficiente [Silva et al. 2024], e a centralização de dados, que impõe entraves de segurança e privacidade [Mendes et al. 2024]. Esses problemas deslocam o foco para a estratégia de treinamento apropriada.

Nesse contexto, o aprendizado federado (*federated learning* – FL) emerge como uma alternativa, permitindo que múltiplos clientes treinem um modelo global, compartilhando apenas parâmetros, e não dados brutos [Grataloup et al. 2024]. Além de evitar a centralização, o FL permite que novos clientes iniciem modelos locais a partir de uma representação pré-treinada, mitigando o *cold start* [Mendes et al. 2024]. Sua aplicação à previsão de demanda, porém, impõe desafios de seleção de participantes [Ramos et al. 2026], de estratégia de agregação [Ali et al. 2025] e de eficiência na comunicação cliente-servidor [Wang et al. 2023].

Dado que operadores de redes de distribuição gerenciam centenas de milhares de clientes, a seleção dos participantes dos modelos globais é uma decisão crítica. Mostra-se então necessário restringir a participação aos clientes mais representativos dos padrões de consumo, sem que haja degradação do desempenho. Reconhecendo essa necessidade, este trabalho propõe uma estratégia de seleção baseada em *clusters* dinâmicos [Bandoria et al. 2026], situada na linha de FL com agrupamento (*Clustered Federated Learning* – CFL). A abordagem revisa periodicamente os *clusters*, define o cliente representativo de cada um e os ranqueia considerando a eficácia *intra-cluster*, o desempenho coletivo e a equidade de seleção.

As principais contribuições desta pesquisa são:

- Uma topologia de *cluster* auto-ajustável, revisada em tempo de execução por operações de fusão e criação a partir da similaridade entre os modelos-*cluster*;
- Mitigar o viés de seleção por meio de um índice de reputação que combina a qualidade do modelo local e um bônus de exploração para clientes e *clusters*;

- Avaliação desses mecanismos na previsão de demanda elétrica residencial, delimitando em que condições os modelos-*cluster* superam um modelo global único.

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta trabalhos relacionados; a Seção 3 descreve o método proposto; a Seção 4 detalha o estudo de caso; a Seção 5 demonstra os resultados; e a Seção 6 discute os resultados e diretrizes futuras.

2. Revisão da Literatura

Para viabilizar essas previsões em cenários de larga escala sem comprometer a privacidade, o artigo de [Ramos et al. 2026] exibe uma abordagem consistente na utilização dos espaços latentes gerados de modelos variacionais (TimeVAE) para seleção de clientes representativos. Essa estratégia permite agrupar perfis de consumo e reduzir custos de comunicação sem a necessidade de compartilhar dados brutos, permitindo a escalabilidade de sistemas de distribuição de energia.

Quanto ao estado da arte, a revisão de [Fu et al. 2023] investiga os desafios de uma seleção eficiente em FL e ressalta a ponderação entre a seleção priorizada e a diversificação feita durante a seleção. Esse compromisso é destacado como um dos principais questionamentos quando a seleção é formulada como uma tarefa de otimização e aprendizado ativo, perspectiva também explorada no trabalho de [Ahmed et al. 2025].

Para exemplificar a seleção como tarefa de otimização, [Wu et al. 2025] propõem uma estratégia baseada em algoritmo genético, cuja função objetivo penaliza as soluções que violam o tempo limite estipulado para cada rodada. Apesar de maximizar o número de participantes, a proposta introduz um elevado custo computacional devido ao processo iterativo da meta-heurística.

Na mesma linha, [Moyeen et al. 2025] rotulam a qualidade de cada cliente a partir dos parâmetros de *hardware* e do desempenho do treinamento local, como contribuição, tempo de convergência e variação dos pesos do modelo. A partir disso, os autores utilizam aprendizado por reforço (*reinforcement learning* – RL) com o objetivo de maximizar a função de reputação modelada. Um dos principais problemas de sua proposta é a presença de clientes com poucos dados, o que pode enviesar o modelo global.

Já [Rashid et al. 2025] avaliam o desempenho do modelo global em um conjunto de dados mantido no servidor, antes e após a contribuição dos modelos locais, ranqueando os clientes por aproximações de *Shapley value*[1] para contornar o custo computacional do cálculo exato. Clientes com baixa reputação perdem prioridade e, quando escolhidos, recebem baixo peso na agregação. Para incentivar a exploração, clientes pouco selecionados recebem bônus iniciais de reputação, evitando que o modelo seja enviesado. Na prática, essa lógica impede que perfis já amplamente explorados sejam agregados repetidamente, mitigando redundâncias e ampliando a diversidade de conhecimento do modelo global. O mecanismo é valioso para a promoção da equidade durante a seleção.

Os trabalhos anteriores tratam o agrupamento como etapa auxiliar da seleção. Uma linha distinta, denominada aprendizado federado com agrupamento (*clustered federated learning* – CFL), assume o agrupamento como elemento estrutural. Em vez

[1] A métrica *Shapley value* associa a cada *input* de modelos de previsão um peso que simboliza o quanto aquela variável foi decisiva para uma determinada previsão. Ela é complexa de calcular e, por isso, é muitas vezes aproximada utilizando outras técnicas.

de um único modelo global, mantém-se um modelo por *cluster*, especializado em um subconjunto de perfis. O FedGroup [Duan et al. 2021], por exemplo, particiona os clientes segundo a similaridade das direções de atualização de seus modelos locais. Mais próximo do domínio deste trabalho, [Gholizadeh and Musilek 2022] emprega *clustering* para previsão de carga elétrica, realizando o agrupamento dos clientes por meio dos hiperparâmetros ajustados a cada rodada, aliado à filtragem de participantes prejudiciais, com comportamento monitorado pela perda do treinamento local. Embora essa abordagem preserve a estabilidade do modelo global, ela pode introduzir viés de seleção ao excluir perfis de consumo que divergem da média dos *clusters* estabelecidos.

Em síntese, a literatura evidencia as vantagens do *clustering* dinâmico, do ranqueamento de clientes e da promoção da equidade, mas deixa duas lacunas. Primeiro, as propostas de CFL revisadas mantêm a topologia de *clusters* essencialmente estática após a formação, e o agrupamento de [Gholizadeh and Musilek 2022], embora identifique clientes representativos, não incorpora essa informação ao critério de seleção. Segundo, os mecanismos de equidade de [Rashid et al. 2025] atuam apenas no nível do cliente, sem considerar *clusters* inteiros que deixam de ser amostrados.

Este trabalho aborda essas lacunas com uma estratégia que opera em duas frentes: *a* uma topologia autoajustável, que funde e cria *clusters* em tempo de execução; e (*b*) um índice multifatorial de reputação com bônus de exploração nos níveis de cliente e de *cluster*. A contribuição não está em usar *clustering* ou reputação em FL isoladamente, mas na integração dos dois mecanismos e na sua avaliação na previsão de demanda residencial. A hipótese central é que modelos-*cluster*, especializados em subconjuntos homogêneos de perfis, alcançam erro médio inferior ao de um modelo global único treinado sobre todos os clientes, o que permite quantificar o compromisso entre especialização e generalização em FL.

3. Metodologia

A estratégia de seleção proposta neste trabalho substitui o uso de um único modelo global pelo uso de vários, cada um associado a um *cluster* e chamado de modelo-*cluster*. Em paralelo, a rotina emprega uma função de reputação associada a cada cliente para priorizar os mais representativos de cada *cluster*. O algoritmo atualiza dinamicamente a formação dos *clusters* e os valores de reputação, com o objetivo de definir quais clientes devem ser treinados e agregados nas rodadas seguintes. A finalidade é mitigar a degradação dos modelos-*cluster* causada pela agregação de clientes sem verificação de qualidade, além de reduzir o custo atrelado ao treinamento de múltiplos clientes.

3.1. Clusterização e Reputação

No contexto de FL, o agrupamento deve ser guiado pelos parâmetros dos modelos e não pelos dados brutos, que permanecem nos clientes. A abordagem responde ao problema de elevado erro individual, substituindo um único modelo global por múltiplos modelos, cada um representativo de um *cluster*. A rotina inicia-se com a formação arbitrária de *clusters* associados a modelos iniciais aleatórios e, por algumas rodadas, com a agregação direta de cada *cluster* com o seu cliente, para que os *clusters* tomem as formas destes. Em seguida, os participantes e seus modelos são atualizados iterativamente até que se formem *clusters* capazes de representar os distintos perfis de demanda presentes nos dados, por dois mecanismos: *merge* e criação de novos *clusters*.

O *merge* consiste na união de dois *clusters* distintos cujos modelos apresentem similaridade acima de um limiar τ_{sup}. O servidor agrega os dois modelos com o mesmo operador de agregação empregado nas rodadas federadas (FedAvg) e consolida em um único conjunto os clientes de ambos os *clusters*. A criação de um novo *cluster* ocorre no servidor quando o modelo treinado por um cliente selecionado apresenta similaridade com seu modelo-*cluster* atual inferior ao limiar τ_{inf}. Nesse caso, o modelo do cliente é comparado aos demais modelo-*clusters*. Se a similaridade com o melhor candidato exceder τ_{sup}, o cliente é realocado para ele. Caso contrário, um novo *cluster* unitário é criado. Neste trabalho, os limiares superiores τ_{sup} e inferiores τ_{inf} foram definidos manualmente por meio de testes da aplicação. A medida de similaridade adotada é o cosseno entre os valores de pesos dos modelos, reescalado para o intervalo $[0, 1]$, em que valores maiores indicam maior semelhança.

Os limiares são definidos manualmente, por inspeção do número de *clusters* resultante em execuções preliminares. Valores muito altos provocam fusões sucessivas e colapsam a topologia, enquanto valores muito baixos produzem fragmentação em *clusters* unitários. Não foi conduzida busca sistemática nem análise de sensibilidade.

A seleção de clientes constitui uma etapa crucial da construção dos *clusters*. Para tal, implementou-se uma seleção por reputação: um valor numérico associado a cada cliente que representa a relevância de sua participação. A cada rodada, o valor escolhe k dentre os *clusters* existente e, em cada um deles, seleciona o cliente de maior reputação como representante, formando o ranqueamento intra-*cluster*, e não global.

Inicialmente, a reputação resume-se aos metadados M associados ao poder computacional e de dados de cada cliente: tamanho local, qualidade de conexão e capacidade de processamento. Como o estudo de caso simula clientes com o mesmo perfil computacional, adota-se $M = 0$ para todos, de modo que a primeira seleção é equiprovável e as diferenças de reputação emergem apenas do treinamento. A partir da primeira rodada, o valor é atualizado pela qualidade do modelo treinado e pelos bônus de exploração.

Clientes que treinaram modelos com baixos erros são interessantes de selecionados novamente, pois demonstram boa capacidade preditiva sobre seus próprios dados, sendo essa a definição de qualidade de um cliente adotada nesse trabalho. No entanto, usar essa métrica como fator principal na seleção causa enviesamento: clientes de menor erro raramente mudam, pois tal condição é fortemente influenciada pela qualidade dos dados, invariável durante a simulação.

Para mitigar isso, emprega-se o bônus de exploração. que acresce valor aos clientes menos selecionados e aos *clusters* subutilizados, incentivando o servidor a agregar participantes cujos dados ainda não estejam representados. Tanto o tempo limite de inatividade quanto o grau do bônus são parâmetros ajustáveis. Matematicamente, o índice de reputação é descrito por:

$$R^{(t)} = Q^{(t)} + \alpha \cdot E_u^{(t)} + \beta \cdot E_g^{(t)} \quad (R^{(0)} = M), \tag{1}$$

onde $R^{(t)}$ é a reputação do cliente na rodada t; M os metadados iniciais; $Q^{(t)}$ é a qualidade do modelo treinado na rodada t, calculado segundo uma função inversa do erro; $E_u^{(t)}$ e $E_g^{(t)}$ são contadores de rodadas consecutivas em que o cliente e seu *cluster* não foram selecionados; e $\alpha = 0{,}2$ e $\beta = 0{,}4$ são os pesos de controle de exploração.

O Algoritmo 1 descreve a interação desses processos com a rotina tradicional de FL, cobrindo a clusterização, a construção da reputação e as principais etapas da estratégia de seleção. As seções seguintes detalham as etapas de pré-processamento, treinamento e avaliação de desempenho dos modelos utilizando o método proposto.

Algoritmo 1: Rotina de treinamento do método.

Entrada: n (número inicial de *clusters*), k (número de *clusters* por rodada), τ_{inf} (limiar de diferença), τ_{sup} (limiar de similaridade)

```
Inicializa n clusters arbitrários;
Inicializa reputações dos clientes;
para algumas rodadas iniciais faça
    Treine e agregue diretamente cada modelo recebido e cluster;
fim para
para cada rodada federada restante faça
    Seleciona clientes representativos para os k clusters atuais (um por
     cluster);
    Clientes realizam o treinamento local;
    Modelos locais são enviados ao servidor;
    para cada modelo cluster recebido Mc_rec faça
        se similaridade(Mc_rec, Mc_relativo) < τ_inf então
            alocado ← falso;
            para cada modelo de cluster Mc_i faça
                se similaridade(Mc_i, Mc_rec) > τ_sup então
                    Cliente entra no cluster i;
                    alocado ← verdadeiro;
                    interromper;
                fim-se
            fim-para
            se não alocado então
                Cliente entra em um novo cluster unitário;
            fim-se
        fim-se
    fim-para
    para cada par de clusters distintos Mc_i e Mc_j faça
        se similaridade(Mc_i, Mc_j) > τ_sup então
            Grupos são unidos (merge);
        fim-se
    fim-para
    Os mapas de clusters e clientes são atualizados;
    Atualiza as reputações através da equação de R^(t);
fim-para
```

3.2. Pré-Processamento

Com os *clusters* formados e os clientes representativos selecionados, os dados de cada cliente são organizados para uso pelos modelos preditivos. O pré-processamento é subdividido nas seguintes operações:

1. **Limpeza de dados**: eliminam-se colunas sem validade semântica, como índices e identificadores;
2. **Tratamento de valores não-numéricos**: substituem-se por parâmetros constantes, pois apenas dados numéricos são processados no treinamento.
3. **Divisão dos dados**: separam-se amostras de treinamento e teste, permitindo a validação objetiva de cada modelo local e dos modelo-*cluster*.
4. **Padronização de escala**: normalizam-se os dados para as mesmas escalas, uma vez que escalas heterogêneas podem enviesar os modelos. Utilizou-se normalização mínimo-máximo com intervalo $[-1, 1]$.
5. **Janelamento temporal**: geram-se os *timelags* (valores históricos usados como novas *features*) por *shift* temporal, etapa fundamental porque grande parte da informação de consumo *day-ahead* reside histórico.

Os dados pré-processados são salvos localmente em cada cliente em formatos de arquivo otimizados, acelerando o acesso local durante as sucessivas rodadas. Os dados brutos não deixam o cliente em nenhuma etapa.

3.3. Pós-Processamento

Esta etapa corresponde ao treinamento local, à agregação e à avaliação dos modelos. O treinamento local segue a abordagem convencional de aprendizado profundo, em que múltiplas arquiteturas de redes neurais podem ser selecionadas para otimizar pesos e vieses utilizados na predição.

Após cada treinamento local, os modelos dos clientes são avaliados e os resultados retornados ao servidor para atualização de reputação e validação global. A métrica adotada é o Erro Médio Quadrático (*Mean Squared Error* – MSE). Os dados utilizados na avaliação nunca foram vistos durante o treinamento, pois foram separados na etapa de pré-processamento, evitando viés de avaliação.

Após o servidor receber os modelos treinados de cada cliente selecionado, executa-se o processo de agregação com os *modelos-cluster*. A agregação consolida o conhecimento de múltiplos modelos de aprendizado de máquina. Diversos algoritmos podem ser empregados; este trabalho utiliza o FedAvg pela simplicidade e para permitir maior foco no impacto da seleção de clientes. Este algoritmo agrega modelos por meio de uma média dos pesos ponderada pelo número de amostras locais, operador também empregado na fusão de dois *clusters*. A seção seguinte apresenta o estudo de caso e os parâmetros utilizados nas simulações. A finalidade é a demonstração dos resultados pela aplicação da metodologia de seleção proposta.

4. Estudo de Caso

Os dados provêm da base de dados Pecan Street Inc. [Street 2024], coleção aberta com informações detalhadas de consumo de eletricidade, água e gás de residências nos Estados Unidos. Utilizou-se a base da cidade de Austin, Texas, com resolução temporal de 15 minutos. O modelo recebe como entrada o histórico do dia anterior, isto é, um *timelag* de 96 observações consecutivas.

O objetivo é prever os valores *day-ahead* da demanda elétrica residencial, representada pela variável `grid`. A base contém ainda o consumo de múltiplos eletrodomésticos por residência, cerca de 40 aparelhos, com grande variação entre domicílios.

Adicionalmente, os dados incluem a variável `solar`, que representa a geração de energia solar (tratada como demanda negativa) para residências equipadas com painéis e as variáveis técnicas `legv1` e `legv2`, correspondentes às tensões de linha.

Todas essas variáveis compõem o vetor de entradas do modelo, permitindo que diferentes residências contribuam com informações específicas de seus padrões de consumo para a construção de modelos-*cluster* especializados. Esta diversidade de características possibilita que o método capture as particularidades de cada cliente, resultando em modelos preditivos mais precisos e bem adaptados aos contextos locais.

5. Resultados

Os resultados apresentados foram obtidos com Python 3.13 e PyTorch 2.8.0, com suporte ao *CUDA Development Kit* em uma placa NVIDIA A30 (24GB HBM2e). A arquitetura implementada foi a *Long Short-Term Memory* (LSTM).

Os hiperparâmetros foram selecionados conforme padrões estabelecidos na literatura e limitações tecnológicas de teste. Utilizou-se: ($i.$) 15 rodadas federadas; ($ii.$) amostra de 6 clientes, inicializada como $n = 6$ *clusters* unitários; ($iii.$) $k = 3$ *clusters* amostrados por rodada ($iv.$) similaridade de cosseno reescalada[2]; ($v.$) limiar de similaridade $\tau_{sup} = 0{,}875$; ($vi.$) limiar de diferença $\tau_{inf} = 0{,}6$; ($vii.$) taxa de aprendizado de 0,001; ($viii.$) 20 épocas locais de treinamento; ($ix.$) tamanho de lote de 1024; ($x.$) metadados iniciais $M = 0$ para todos os clientes.

O desempenho do método é comparado ao de uma linha de base composta por um modelo global único, treinado com FedAvg sobre as mesmas rodadas e avaliado no mesmo conjunto de teste. Essa linha de base é federada. Os erros medidos pelo MSE entre a previsão e o valor real correspondente são exibidos na Tabela 1, onde as células vazias indicam *clusters* que deixaram de existir após a função.

Tabela 1. Erro médio dos *clusters* ativos e do modelo global ao longo das rodadas.

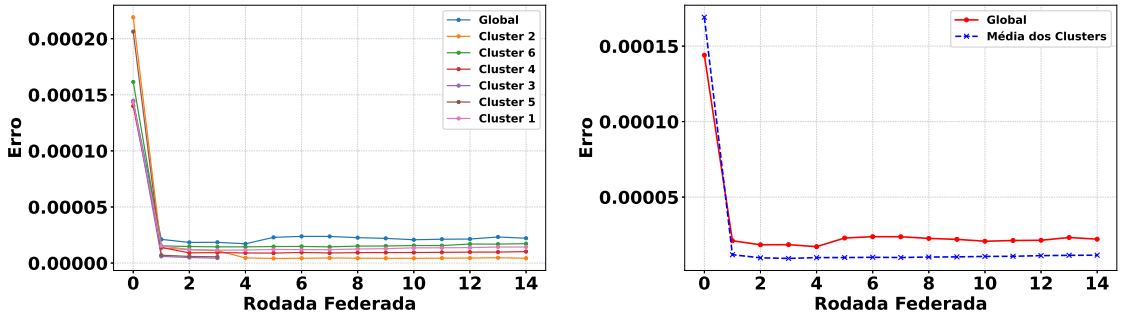
Rod.	*Cluster* 1	*Cluster* 2	*Cluster* 3	*Cluster* 4	*Cluster* 5	*Cluster* 6	$\lvert C_t \rvert$	Média	Global
0	0,000219	0,000161	0,000140	0,000145	0,000206	0,000143	6	0,000169	0,000144
1	0,000014	0,000015	0,000014	0,000006	0,000007	0,000015	6	0,000012	0,000021
2	0,000012	0,000015	0,000009	0,000005	0,000006	0,000012	6	0,000010	0,000018
3	0,000011	0,000014	0,000009	0,000004	0,000006	0,000011	6	0,000009	0,000018
4	0,000005	0,000014	0,000009	-	-	0,000012	4	0,000010	0,000017
5	0,000004	0,000015	0,000009	-	-	0,000012	4	0,000010	0,000023
6	0,000004	0,000015	0,000009	-	-	0,000012	4	0,000010	0,000024
7	0,000005	0,000014	0,000009	-	-	0,000012	4	0,000010	0,000024
8	0,000004	0,000015	0,000009	-	-	0,000013	4	0,000010	0,000023
9	0,000004	0,000015	0,000009	-	-	0,000013	4	0,000010	0,000022
10	0,000004	0,000016	0,000009	-	-	0,000014	4	0,000011	0,000021
11	0,000004	0,000016	0,000010	-	-	0,000014	4	0,000011	0,000021
12	0,000004	0,000017	0,000010	-	-	0,000014	4	0,000011	0,000021
13	0,000005	0,000017	0,000010	-	-	0,000014	4	0,000011	0,000023
14	0,000004	0,000017	0,000010	-	-	0,000014	4	0,000012	0,000022

[2]A similaridade de cosseno entre os vetores de pesos de dois modelos é reescalada para o intervalo $[0, 1]$, em que valores maiores indicam maior semelhança.

A Tabela 1 ilustra que determinados *clusters* passam por processos de fusão, deixando de participar das rodadas seguintes. Observa-se que os *clusters* 4 e 5 foram absorvidos após a terceira rodada, visto que eram suficientemente similares para tal união. Além disso, para que a comparação entre os modelos dos *clusters* e o modelo global seja consistente, é necessário uma métrica que não seja afetada pela quantidade de agrupamentos. Uma vez que a soma dos erros falha nesse aspecto, optou-se por utilizar a média ponderada do MSE dos *clusters* ativos em cada rodada.

Desconsiderando-se a rodada 0, anterior a qualquer agregação, a média da ordem de grandeza do erro nas rodadas de 1 a 14 é de $10^{-4,676}$ para os modelos dos *clusters* e de $10^{-4,530}$ para o modelo global, valores obtidos por meio do logaritmo do erro médio correspondente. Desse modo, os modelos dos *clusters* apresentam um erro médio inferior, com uma diferença de aproximadamente $0{,}146$ ordens de magnitude. Esse resultado corrobora a hipótese central de que a especialização por *cluster* reduz o erro agregado, preservando a convergência dos modelos. Outra vantagem da abordagem baseada em *clusters* é a sua velocidade de convergência, visto que seus erros se estabilizam em patamares baixos mais rapidamente do que os do modelo global, o qual exigiria rodadas federadas adicionais para alcançar a estabilidade.

Adicionalmente, observa-se que a técnica de *clustering* é fortemente favorecida quando há equilíbrio entre τ_{sup} e τ_{inf}: valores muito altos causam fusão excessiva de *clusters*, enquanto valores muito baixos produzem fragmentação. Este equilíbrio permite que o número de *clusters* se mantenha em faixa ótima para a representação dos padrões de consumo. A Figura 1 resume graficamente a evolução dos erros da Tabela 1.



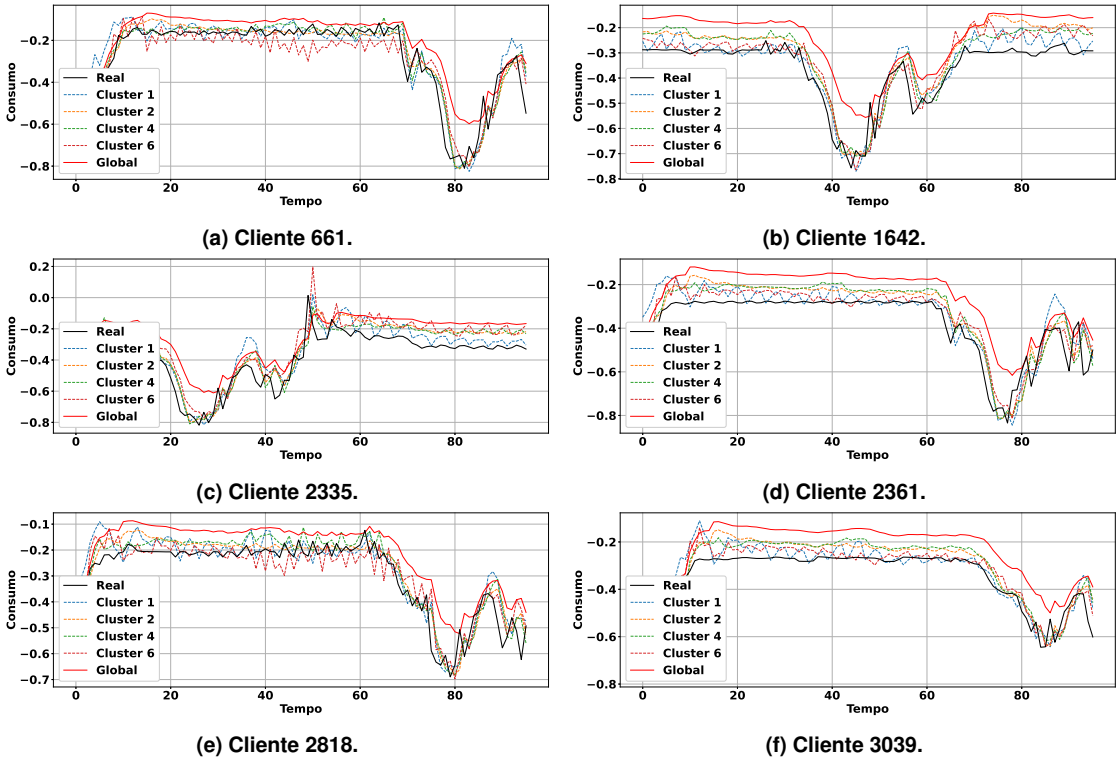
(a) Modelo global *vs. clusters*. **(b) Média dos *clusters vs.* modelo global.**

Figura 1. Comparação entre os erros dos *clusters* e global.

A Figura 1 demonstra que, após a queda inicial observada em todos os modelos, as curvas dos *clusters* se estabilizam abaixo da curva global e apresentam menor dispersão entre si. Esse comportamento é consistente com a premissa de que *clusters* homogêneos são mais fáceis de ajustar do que o conjunto completo de perfis. Contudo, ressalta-se que o experimento não isola a contribuição da reputação, uma vez que não foi executada uma configuração de controle com *clustering* e seleção aleatória.

Por fim, a Figura 2 ilustra as predições para seis clientes, em valores normalizados, com o intuito de confrontar o desempenho do modelo global em relação ao modelo do *cluster* designado a cada cliente ao término do treinamento. A inspeção visual corrobora a Tabela 1. De modo geral, o modelo do *cluster* correspondente adere com maior precisão à curva real, ao passo que o modelo global apresenta desvios mais acentuados devido à

sua tentativa de generalizar múltiplos perfis. Casos representativos dessa dinâmica podem ser observados nos clientes 1642 (*Cluster* 1), 661 (*Cluster* 4) e 3039 (*Cluster* 6).



(a) Cliente 661. (b) Cliente 1642.

(c) Cliente 2335. (d) Cliente 2361.

(e) Cliente 2818. (f) Cliente 3039.

Figura 2. Previsões diversas do modelo global e modelos-*cluster*.

6. Conclusão

Este trabalho investigou uma estratégia de seleção de clientes para FL que combina agrupamento dinâmico e reputação com bônus de exploração. No estudo de caso conduzido com a arquitetura LSTM, os modelos dos *clusters* alcançaram um erro médio inferior em cerca de $0{,}146$ ordens de magnitude em comparação a um modelo global único treinado com FedAvg. Isso sustenta a hipótese de que a especialização por *cluster* reduz o erro agregado. Além disso, a análise das previsões individuais mostrou que essa vantagem se transfere para cada curva de consumo, sendo que clientes pertencentes a um *cluster* são melhor representados pelo seu respectivo modelo. Em conjunto, esses resultados delimitam o alcance da proposta, demonstrando que ela é vantajosa quando o objetivo é o desempenho individualizado por perfil, embora exiba sensibilidade aos hiperparâmetros escolhidos.

É importante destacar que a avaliação se restringe a um número limitado de clientes de mesma localidade, poucas rodadas federadas, uma única arquitetura e a uma execução isolada, sem repetição com múltiplas sementes, o que impede afirmações estatísticas sobre a estabilidade. O FedAvg foi o único método de agregação considerado, e os limiares τ_{inf} e τ_{sup} foram definidos manualmente, sem busca sistemática ou análise de sensibilidade, gerando resultados não ideais para a representação dos perfis. Não houve comparação experimental com outros métodos de CFL, nem uma configuração de controle com *clustering* e seleção aleatória, de modo que os ganhos observados não podem

ser atribuídos separadamente ao agrupamento ou à reputação. Pela mesma razão, o bônus de exploração é aqui descrito apenas como um mecanismo projetado para desencorajar a exclusão persistente de clientes e *clusters*, e não como uma melhoria de equidade. A técnica de *clustering* também poderia ser aprimorada utilizando atualizações dos gradientes ao invés dos pesos e viéses do modelo. Por fim, não foram coletadas métricas de tempo de processamento ou de custo de comunicação. Cabe notar que a etapa de fusão compara pares de *clusters* e cresce quadraticamente com o número de *clusters* ativos, o que exige atenção em cenários de larga escala. Tais limitações configuram oportunidades diretas para trabalhos futuros.

Agradecimentos

Os autores agradecem o suporte recebido pela Universidade Federal de Juiz de Fora (UFJF), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Rede Integrada de Pesquisa em Alta Velocidade (RePesq).

Referências

Ahmed, S., Shamim Kaiser, M., Chaki, S., and Shawkat Ali, A. B. M. (2025). Adaptive federated learning with reinforcement learning-based client selection for heterogeneous environments. *IEEE Access*, 13:131671–131695.

Ali, S. S., Ali, M., Bhatti, D. M. S., and Choi, B. J. (2025). Explainable clustered federated learning for solar energy forecasting. *Energies*, 18(9).

Azzalini, D., Flammini, B., Emanuele, C. A., Guadagno, A., Ragaini, E., and Amigoni, F. (2025). An empirical evaluation of deep autoencoders for anomaly detection in the electricity consumption of buildings. *Energy and Buildings*, 327:115069.

Bandoria, L. H., Silva, W. N., Dias, B. H., and Almeida, M. C. (2026). Self-supervised learning for advanced load profile segmentation in distribution networks. *Sustainable Energy, Grids and Networks*, 46:102166.

Casolaro, A., Capone, V., Iannuzzo, G., and Camastra, F. (2023). Deep learning for time series forecasting: Advances and open problems. *Information*, 14(11):598.

Dakheel, F. and Çevik, M. (2025). Optimizing smart grid load forecasting via a hybrid long short-term memory-xgboost framework: Enhancing accuracy, robustness, and energy management. *Energies*, 18(11).

Duan, M., Liu, D., Ji, X., Liu, R., Liang, L., Chen, X., and Tan, Y. (2021). Fedgroup: Efficient federated learning via decomposed similarity-based clustering. In *2021 IEEE Intl Conf on parallel & distributed processing with applications, big data & cloud computing, sustainable computing & communications, social computing & networking*, pages 228–237.

Fu, L., Zhang, H., Gao, G., Zhang, M., and Liu, X. (2023). Client selection in federated learning: Principles, challenges, and opportunities. *IEEE Internet of Things Journal*, 10(24):21811–21819.

Gholizadeh, N. and Musilek, P. (2022). Federated learning with hyperparameter-based clustering for electrical load forecasting. *Internet of Things*, 17:100470.

Grataloup, A., Jonas, S., and Meyer, A. (2024). A review of federated learning in renewable energy applications: Potential, challenges, and future directions. *Energy and AI*, 17:100375.

He, M. (2024). Spatial-temporal prediction of electric vehicle charging load based on multi-time step semi-dynamic traffic equilibrium. In *2024 5th International Conference on Clean Energy and Electric Power Engineering (ICCEPE)*, pages 1243–1248. IEEE.

Humiston, C., Cetin, M., and Queiroz, A. R. d. (2026). Evaluating battery degradation models in rolling-horizon bess arbitrage optimization. *Energies*, 19(4):1056.

Mendes, N., Mendes, J., Mohammadi, J., and Moura, P. (2024). Federated learning framework for prediction of net energy demand in transactive energy communities. *Sustainable Energy, Grids and Networks*, 40:101522.

Moyeen, M., Kaur, K., Agarwal, A., Manzano, S. R., Zaman, M., and Goel, N. (2025). Fed-reputed: Reputation-aware client selection in hierarchical federated learning for consumer electronics. *IEEE Transactions on Consumer Electronics*.

Navidi, T., El Gamal, A., and Rajagopal, R. (2023). Coordinating distributed energy resources for reliability can significantly reduce future distribution grid upgrades and peak load. *Joule*, 7(8):1769–1792.

Ramos, P. V. B., Silva, W. N., Bandória, L. H., Dias, B. H., Villela, S. M., and Morais, H. (2026). Federated electricity demand forecasting: A client selection mechanism based on timevae latent representations. *Applied Energy*, 419:128064.

Ramos, P. V. B., Villela, S. M., Silva, W. N., and Dias, B. H. (2023). Residential energy consumption forecasting using deep learning models. *Applied Energy*, 350:121705.

Rashid, M. M., Xiang, Y., Uddin, M. P., Tang, J., Sood, K., and Gao, L. (2025). Trustworthy and fair federated learning via reputation-based consensus and adaptive incentives. *Trans. Info. For. Sec.*, 20:2868–2882.

Silva, W. N., Ramos, P. V. B., Dias, B. H., Almeida, M. C., Tenório, L. H., Oliveira, L. W., and Villela, S. M. (2024). Forecasting demand in data-limited environments: a benchmarking study. In *Congresso Brasileiro de Automática-CBA*, volume 4.

Street, P. (2024). Pecan street dataport. Acessado em: 25 mar. 2024.

Thwe, M. M., Ştefanov, A., Rajkumar, V. S., and Palensky, P. (2025). Digital twins for power systems: Review of current practices, requirements, enabling technologies, data federation and challenges. *IEEE Access*.

Wang, Z., Wen, M., Xu, Y., Zhou, Y., Wang, J. H., and Zhang, L. (2023). Communication compression techniques in distributed deep learning: A survey. *Journal of Systems Architecture*, 142.

Wu, J., Ji, H., Yi, J., and Liu, L. (2025). Optimizing client selection in federated learning base on genetic algorithm. *Cluster Computing*, 28(6):400.