

Otimização de Modelos de Aprendizado Federado com Redes Definidas por Software

Cândido Leandro de Queiroga Bisneto



CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa, 2025

Cândido Leandro de Queiroga Bisneto

Otimização de Modelos de Aprendizado Federado com Redes Definidas por Software

Monografia apresentada ao curso Ciência da Computação
do Centro de Informática, da Universidade Federal da Paraíba,
como requisito para a obtenção do grau de Bacharel em Ciência da Computação

Orientador: Fernando Menezes Matos

Dezembro de 2025



CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

Trabalho de Conclusão de Curso de Ciência da Computação intitulado ***Otimização de Modelos de Aprendizado Federado com Redes Definidas por Software*** de autoria de Cândido Leandro de Queiroga Bisneto, aprovada pela banca examinadora constituída pelos seguintes professores:

Prof. Dr. Fernando Menezes Matos
Universidade Federal da Paraíba

Prof. [Nome do Professor B]
Universidade Federal da Paraíba

Prof. [Nome do Professor C]
Universidade Federal da Paraíba

Coordenador(a) do Curso de Ciência da Computação
[Nome do Coordenador]
CI/UFPB

João Pessoa, Dezembro de 2025

DEDICATÓRIA

[Texto da dedicatória - opcional]

AGRADECIMENTOS

[Texto dos agradecimentos]

RESUMO

[Inserir resumo em português - um único parágrafo informativo contendo: objetivos do trabalho, justificativa e resultados alcançados. Máximo 200 palavras.]

Palavras-chave: Aprendizado Federado, Redes Definidas por Software, XGBoost, LightGBM, CatBoost, Flower Framework, Privacidade de Dados.

ABSTRACT

[Insert abstract in English - one informative paragraph containing: work objectives, justification, and results achieved. Maximum 200 words.]

Keywords: Federated Learning, Software-Defined Networking, XGBoost, LightGBM, CatBoost, Flower Framework, Data Privacy.

LISTA DE FIGURAS

LISTA DE TABELAS

1	Tecnologias utilizadas no projeto	20
2	Metadados do dataset	21
3	Especificações do ambiente computacional	22

LISTA DE ABREVIATURAS

API - Application Programming Interface

AUC - Area Under the Curve

CART - Classification and Regression Trees

CPU - Central Processing Unit

DP - Differential Privacy

FL - Federated Learning

GBDT - Gradient Boosting Decision Trees

GPU - Graphics Processing Unit

gRPC - Google Remote Procedure Call

HTTP - Hypertext Transfer Protocol

IID - Independent and Identically Distributed

LGPD - Lei Geral de Proteção de Dados

ML - Machine Learning

non-IID - Non-Independent and Identically Distributed

QoS - Quality of Service

ROC - Receiver Operating Characteristic

SDN - Software-Defined Networking

TCC - Trabalho de Conclusão de Curso

TCP - Transmission Control Protocol

UFPB - Universidade Federal da Paraíba

UDP - User Datagram Protocol

Sumário

1	INTRODUÇÃO	17
1.1	Motivação	17
1.2	Objetivo	17
1.3	Estrutura da monografia	17
2	REVISÃO DE LITERATURA	18
2.1	Trabalhos Relacionados sobre Aprendizado Federado	18
2.2	Trabalhos sobre Modelos Baseados em Árvores	18
2.3	Trabalhos sobre SDN e Otimização de Rede	18
2.4	Análise Comparativa dos Trabalhos	18
3	FUNDAMENTOS TEÓRICOS	19
3.1	Aprendizado Federado	19
3.1.1	Definição e Características	19
3.1.2	Estratégias de Agregação	19
3.2	Modelos Baseados em Árvores de Decisão	19
3.2.1	XGBoost	19
3.2.2	LightGBM	19
3.2.3	CatBoost	19
3.3	Redes Definidas por Software (SDN)	19
3.3.1	Arquitetura SDN	19
3.3.2	OpenFlow	19
3.3.3	Aplicações em Aprendizado de Máquina	19
3.4	Framework Flower	19
3.4.1	Arquitetura do Flower	19
3.4.2	Implementação de Clientes e Servidores	19
4	METODOLOGIA	20
4.1	Visão Geral da Solução Proposta	20

4.2	Arquitetura do Sistema	20
4.2.1	Componentes do Sistema	20
4.2.2	Fluxo de Comunicação	20
4.3	Implementação dos Modelos Federados	20
4.3.1	Adaptação de XGBoost para FL	20
4.3.2	Adaptação de LightGBM para FL	20
4.3.3	Adaptação de CatBoost para FL	20
4.4	Integração com SDN	20
4.4.1	Configuração da Rede SDN	20
4.4.2	Políticas de QoS	20
4.4.3	Monitoramento de Tráfego	20
4.5	Tecnologias Utilizadas	20
5	DATASET E PREPARAÇÃO DOS DADOS	21
5.1	Seleção do Dataset	21
5.2	Descrição das Features	21
5.3	Análise Exploratória dos Dados	21
5.4	Pré-processamento	21
5.4.1	Tratamento de Valores Ausentes	21
5.4.2	Normalização e Padronização	21
5.4.3	Codificação de Variáveis Categóricas	21
5.5	Particionamento dos Dados	21
5.5.1	Particionamento IID	21
5.5.2	Particionamento non-IID	21
6	CONFIGURAÇÃO EXPERIMENTAL	22
6.1	Ambiente Computacional	22
6.2	Hiperparâmetros dos Modelos	22
6.3	Configuração do Aprendizado Federado	22
6.4	Experimentos Realizados	22

6.5	Métricas de Avaliação	22
7	RESULTADOS	23
7.1	Resultados em Cenário IID	23
7.1.1	Comparação de Modelos	23
7.1.2	Comparação de Estratégias	23
7.2	Resultados em Cenário non-IID	23
7.2.1	Comparação de Modelos	23
7.2.2	Impacto da Heterogeneidade	23
7.3	Análise de Convergência	23
7.4	Impacto da Integração SDN	23
7.5	Análise Estatística	23
7.5.1	Teste de Friedman	23
7.5.2	Teste Post-hoc de Nemenyi	23
8	DISCUSSÃO	24
8.1	Interpretação dos Resultados	24
8.2	Comparação com Estado da Arte	24
8.3	Limitações do Trabalho	24
8.4	Implicações Práticas	24
9	CONCLUSÕES E TRABALHOS FUTUROS	25
9.1	Conclusões	25
9.2	Contribuições	25
9.3	Trabalhos Futuros	25
	REFERÊNCIAS	25

1 INTRODUÇÃO

1.1 Motivação

1.2 Objetivo

1.3 Estrutura da monografia

Esta monografia está dividida da seguinte maneira: na segunda seção é feita a revisão da literatura sobre aprendizado federado e redes SDN. Na terceira seção são apresentados os fundamentos teóricos necessários para compreensão do trabalho. Na quarta seção é apresentada a metodologia utilizada no desenvolvimento do trabalho. Na quinta seção o dataset utilizado é descrito em detalhes. Na sexta seção a configuração experimental é apresentada. Na sétima seção os resultados obtidos são apresentados e analisados. Na oitava seção é feita a discussão dos resultados. Na nona seção é feita a conclusão deste trabalho e trabalhos futuros são propostos.

2 REVISÃO DE LITERATURA

2.1 Trabalhos Relacionados sobre Aprendizado Federado

2.2 Trabalhos sobre Modelos Baseados em Árvores

2.3 Trabalhos sobre SDN e Otimização de Rede

2.4 Análise Comparativa dos Trabalhos

3 FUNDAMENTOS TEÓRICOS

3.1 Aprendizado Federado

3.1.1 Definição e Características

3.1.2 Estratégias de Agregação

3.2 Modelos Baseados em Árvores de Decisão

3.2.1 XGBoost

3.2.2 LightGBM

3.2.3 CatBoost

3.3 Redes Definidas por Software (SDN)

3.3.1 Arquitetura SDN

3.3.2 OpenFlow

3.3.3 Aplicações em Aprendizado de Máquina

3.4 Framework Flower

3.4.1 Arquitetura do Flower

3.4.2 Implementação de Clientes e Servidores

4 METODOLOGIA

4.1 Visão Geral da Solução Proposta

4.2 Arquitetura do Sistema

4.2.1 Componentes do Sistema

4.2.2 Fluxo de Comunicação

4.3 Implementação dos Modelos Federados

4.3.1 Adaptação de XGBoost para FL

4.3.2 Adaptação de LightGBM para FL

4.3.3 Adaptação de CatBoost para FL

4.4 Integração com SDN

4.4.1 Configuração da Rede SDN

4.4.2 Políticas de QoS

4.4.3 Monitoramento de Tráfego

4.5 Tecnologias Utilizadas

Componente	Tecnologia
Framework FL	Flower 1.6+
Modelos	XGBoost, LightGBM, CatBoost
Linguagem	Python 3.9+
SDN Controller	[A definir]
Comunicação	gRPC

Tabela 1: Tecnologias utilizadas no projeto

5 DATASET E PREPARAÇÃO DOS DADOS

5.1 Seleção do Dataset

Característica	Valor
Nome	[Nome do dataset]
Fonte	[URL/Repositório]
Número de amostras	[N]
Número de features	[M]
Classes	[Lista de classes]
Tipo de problema	Classificação binária/multiclasse

Tabela 2: Metadados do dataset

5.2 Descrição das Features

5.3 Análise Exploratória dos Dados

5.4 Pré-processamento

5.4.1 Tratamento de Valores Ausentes

5.4.2 Normalização e Padronização

5.4.3 Codificação de Variáveis Categóricas

5.5 Particionamento dos Dados

5.5.1 Particionamento IID

5.5.2 Particionamento non-IID

6 CONFIGURAÇÃO EXPERIMENTAL

6.1 Ambiente Computacional

Componente	Especificação
Processador	[Modelo e frequência]
Memória RAM	[Quantidade] GB
GPU	[Modelo] ou N/A
Sistema Operacional	[Nome e versão]
Python	[Versão]

Tabela 3: Especificações do ambiente computacional

6.2 Hiperparâmetros dos Modelos

6.3 Configuração do Aprendizado Federado

6.4 Experimentos Realizados

6.5 Métricas de Avaliação

7 RESULTADOS

7.1 Resultados em Cenário IID

7.1.1 Comparação de Modelos

7.1.2 Comparação de Estratégias

7.2 Resultados em Cenário non-IID

7.2.1 Comparação de Modelos

7.2.2 Impacto da Heterogeneidade

7.3 Análise de Convergência

7.4 Impacto da Integração SDN

7.5 Análise Estatística

7.5.1 Teste de Friedman

7.5.2 Teste Post-hoc de Nemenyi

8 DISCUSSÃO

8.1 Interpretação dos Resultados

8.2 Comparação com Estado da Arte

8.3 Limitações do Trabalho

8.4 Implicações Práticas

9 CONCLUSÕES E TRABALHOS FUTUROS

9.1 Conclusões

9.2 Contribuições

9.3 Trabalhos Futuros

REFERÊNCIAS

- [1] AUTOR, A.B.; COAUTOR, C.D. Título do artigo. **Nome da Revista**, v.10, n.2, p.45-62, 2020.
- [2] AUTOR, Nome. **Título do Livro**. Editora, Cidade, Ano.
- [3] AUTOR, A.; COAUTOR, B. Título do artigo. **Nome da Conferência**, p.100-110, 2021.