



**UNIVERSIDADE ESTADUAL DA PARAÍBA
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA EM SAÚDE**

RENAN MATIAS MOURA

**DESENVOLVIMENTO DE UMA API REST COM MODELO DE MACHINE
LEARNING PARA PREDIÇÃO DE PRÉ-ECLÂMPSIA - PPML**

CAMPINA GRANDE

2025

RENAN MATIAS MOURA

**DESENVOLVIMENTO DE UMA API REST COM MODELO DE MACHINE
LEARNING PARA PREDIÇÃO DE PRÉ-ECLÂMPSIA - PPML**

Projeto encaminhado ao Comitê de Ética em Pesquisa da Universidade Estadual da Paraíba para análise e parecer com a finalidade de cumprir com investigações de estudo para para conclusão do mestrado em Ciência e tecnologia em saude.

Orientador: Prof. Dr. Frederico Moreira Bublitz.

CAMPINA GRANDE

2025

NOME DO ALUNO

TÍTULO DO TRABALHO

Trabalho de Conclusão de Curso apresentado
ao Departamento de XXXXXXXXX do Centro
de XXXXXXXXX da Universidade Estadual da
Paraíba como requisito parcial à obtenção do
título de Licenciado(a) em XXXXXXXX.

Área de concentração: xxxxxxxxxxxx

Aprovado em: _____

BANCA EXAMINADORA

Prof. Dr. xxxxxxxxxxxxxxxx (Orientador)

Universidade Estadual da Paraíba (UEPB)

Profa. Dra. xxxxxxxxxxxxxxxx

Universidade Estadual da Paraíba (UEPB)

Prof. Dr. xxxxxxxxxxxxxxxxx

Universidade XXXXXX (Caso MEMBRO EXTERNO)

Dedicatória.

AGRADECIMENTOS

Escreva aqui os seus agradecimentos...

RESUMO

A pré-eclâmpsia (PE) é uma das causas mais comuns de morbidade e mortalidade materna e perinatal. A doença é caracterizada pelo desenvolvimento de hipertensão após 20 semanas de gestação e pode desencadear uma série de complicações. Em casos mais graves, a condição pode levar à morte, restrição de crescimento fetal e ao parto prematuro, aumentando significativamente o risco de complicações neonatais. Portanto, a detecção precoce e o manejo adequado da doença são essenciais para minimizar desfechos adversos tanto para a mãe quanto para o bebê. A Fetal Medicine Foundation (FMF) desenvolveu e disponibilizou a Calculadora de Risco para a PE, que combina fatores maternos, pressão arterial média e índices de pulsatilidade das artérias uterinas, além de marcadores séricos como o fator de crescimento placentário (PLGF), sendo amplamente reconhecida e utilizada desde o início dos anos 2010, sendo utilizados modelo de regressão logística. Nos últimos anos, técnicas de aprendizado de máquina têm emergido como ferramentas promissoras em aplicações médicas. A aplicação desses modelos baseia-se na análise de grandes volumes de dados clínicos, permitindo a identificação de padrões que podem indicar a doença antes mesmo do aparecimento de sintomas evidentes. Apesar dos avanços, as calculadoras foram validadas utilizando dados que podem não refletir a realidade brasileira, trazendo imprecisão ao diagnóstico, o que pode refletir na qualidade da assistência pré-natal e nos resultados perinatais. Visando esse problema, uma API com modelo de Machine Learning (ML) foi desenvolvida para avaliar a acurácia utilizando dados da nossa população. A ML foi treinada com dados clínicos reais, demonstrando resultados com alta precisão rápidas, podendo otimizar o acompanhamento às gestantes.

Palavras-chave: Aprendizado de máquina. pré-eclâmpsia. doppler das artérias uterinas. triagem do primeiro trimestre.

ABSTRACT

Resumo em inglês

Keywords: Keyword 1. Keyword 2. Keyword 3.

SUMÁRIO

1	INTRODUÇÃO	8
2	OBJETIVOS	11
2.1	Objetivos gerais	11
2.2	Objetivos específicos	11
3	RISCOS E BENEFICIOS	12
3.1	Benefícios	12
3.2	Riscos	12
3.3	Prevenção e minimização dos riscos	12
4	FUNDAMENTAÇÃO TEÓRICA	13
4.1	Pré-Natal	13
4.2	Pré-eclâmpsia	13
4.3	Biomarcadores	14
4.4	Padrões arquitetônicos de softwares	14
4.5	API REST	14
4.6	Machine Learning	15
5	ESTADO DA ARTE	17
6	PPML – API REST FULL	19
7	METODOLOGIA	21
7.1	Análise dos dados	21
	REFERÊNCIAS	22
	APÊNDICES	26

1 INTRODUÇÃO

Atualmente, com a inversão da pirâmide da assistência pré-natal, o foco passou a ser na detecção de pacientes com risco aumentado de desenvolver complicações durante a gestação e oferecer profilaxia, ao invés de esperar a complicações ocorrer e ser tratada. A detecção precoce de riscos gestacionais é fundamental para a saúde da mulher e da criança, principalmente em países como o Brasil, onde a morbimortalidade fetal, neonatal e materna ainda é significativa. Estudos indicam que, embora a cobertura da assistência pré-natal no Brasil seja praticamente universal, a adequação dessa teste assistência permanece baixa, com menos de 10% das gestantes recebendo todos os procedimentos recomendados(Viellas *et al.*, 2014).

Os distúrbios hipertensivos na gravidez estão entre as complicações mais comuns durante a gestação, englobando hipertensão crônica e hipertensão gestacional, sendo a pré-eclâmpsia (PE) a forma mais grave(Garovic *et al.*, 2022). Definida como um distúrbio multifatorial da gravidez, é caracterizada por disfunção placentária e dano vascular sistêmico, que coloca em risco a vida de mães e bebês em todo o mundo. Afetando cerca de 3% a 5% das gestações, chega a ser responsável por mais de 42 mil mortes maternas anuais, com impacto desproporcional em países de baixa e média renda, em que o acesso a cuidados obstétricos de qualidade são na sua maioria limitados(Chappell *et al.*, 2021).

A *Fetal Medicine Foundation* (FMF) foi uma precursora nesta área e desenvolveu uma calculadora de risco para pré-eclâmpsia que combina dados da história materna, pressão arterial média e índices de pulsatilidade das artérias uterinas, além de outros biomarcadores, para estimar o risco de desenvolver pré-eclâmpsia durante a gestação. A calculadora utiliza o teorema de Bayes para combinar o risco prévio derivado de fatores maternos com dados adicionais coletados durante a gestação, permitindo uma avaliação mais precisa do risco de pré-eclâmpsia(Riishede *et al.*, 2023). A utilização dessa ferramenta pode ajudar na identificação precoce de gestantes em risco, permitindo intervenções preventivas e melhorando os resultados maternos e fetais.

Please record the following information and then press Calculate.

Maternal characteristics	
Date of birth	dd/mm/yyyy
Height	cm
Weight	kg
Ethnicity	<input type="button" value="▼"/>
Currently smoking	<input type="radio"/> Yes <input type="radio"/> No
Current pregnancy	
Last menstrual period	dd/mm/yyyy <input type="button" value="..."/>
Vanishing twin	<input type="radio"/> Yes <input type="radio"/> No
Conception method	<input type="button" value="▼"/>
Pregnancy dating	
Method	<input type="button" value="▼"/>
Gestational age	weeks <input type="button" value="▼"/> days
EDD	dd/mm/yyyy
Examination date	01/12/2025

Medical history

- Chronic hypertension Yes No
Diabetes type I Yes No
Diabetes type II Yes No
Mother of the patient had PE Yes No
Systemic lupus erythematosus Yes No
Anti-phospholipid syndrome Yes No

Obstetric history

- Nulliparous (no previous pregnancies at ≥24 wks.)
 Parous (at least one pregnancy at ≥24 weeks)

For more information on the background of this application and interpretation of the results please [click here](#).

This software is based on research carried out by The Fetal Medicine Foundation. Neither the FMF nor any other party involved in the development of this software shall be held liable for results produced using data from unconfirmed sources. Clinical risk assessment requires that the ultrasound measurements are taken by accredited practitioners.

Figura 1 – Fonte: Fetal Medicine Foundation, 2025.

Pesquisas demonstram que modelos de aprendizado de máquina aplicados a dados de gestantes são capazes de classificar gravidez de risco com alta precisão, permitindo intervenções precoces(Silva, 2021), portanto identificar precocemente essas condições permite que os profissionais de saúde adotem medidas preventivas, ofereçam orientação adequada e garantam intervenções oportunas, reduzindo assim complicações durante a gestação e o parto.

A importância de um sistema de avaliação de riscos gestacionais que possa ser utilizado no sistema público tem o potencial de salvar vidas, melhorar a qualidade dos cuidados maternos e promover a saúde das gestantes e de seus bebês em curto, médio e longo prazo. Portanto, a busca por soluções utilizando a Inteligência Artificial e modelos de avaliação de riscos gestacionais adaptados ao contexto brasileiro é um passo crucial na melhoria da saúde materna e perinatal no país. Iniciativas como a aplicação de modelos preditivos baseados em inteligência artificial

têm mostrado eficácia na otimização dos cuidados neonatais e na redução da mortalidade materna(Alcino *et al.*, 2024).

Abordagens com ML para predição de PE (todos os tipos) têm demonstrado alta precisão, alcançando taxas de detecção acima dos 80% e taxa de falsos positivos a 10% superando as limitações de métodos anteriores que não consideravam variações populacionais. A capacidade de processar dados de forma dinâmica e fornecer recomendações baseadas em evidências em tempo real tem o potencial de revolucionar a aplicação clínica desses preditores, permitindo intervenções precoces e personalizadas. No entanto, para que esses modelos sejam amplamente adotados, é necessário superar desafios como a validação em diferentes populações e a integração de biomarcadores como o fator de crescimento placentário (PLGF) em contextos de atenção básica, onde o acesso a esses recursos ainda é limitado(Gil *et al.*, 2024).

Apesar da alta sensibilidade e especificidade apresentada pelos algoritmos disponibilizados pela FMF para prever PE, a validação externa é questionada devido ao fato de terem seus parâmetros calibrados a partir de dados europeus. Apesar de já ter sido validada para gestantes brasileiras, questiona-se se não poderíamos ter uma acurácia maior se utilizássemos dados da nossa população na calibração de seu algoritmo(Rezende *et al.*, 2024).

Portanto desenvolvemos um API treinada e testada, pronta para integração com outros softwares, utilizando um conjunto de informações do pré-natal e consultas ultrassonográficas, sem o uso de biomarcadores. Esses dados são originados do atendimento diario, no qual são armazenadas as informações dos pacientes em um sistema de laudos de ultrassom pré-existente a esse trabalho e um levantamento dos desfechos das gestações que foram acompanhadas na clínica privada EMBRION com sede em Campina Grande-PB.

2 OBJETIVOS

2.1 Objetivos gerais

Melhorar o acesso a ferramentas de predição de pré-eclâmpsia, especialmente em contextos com acesso limitado a biomarcadores, por meio de um sistema integrado a Inteligência Artificial (IA).

2.2 Objetivos específicos

Desenvolver uma API REST com um modelo de *Machine Learning* (ML) treinado com dados maternos e fetais, sem uso de dados de biomarcadores, que seja capaz de predizer a pré-eclâmpsia com uma precisão melhor ou igual ao método da FMF. Como também oferecer uma integração REST com softwares de estabelecimentos de saúde que ofertam o serviço de pré-natal, visando a identificação precoce de gestantes com maior risco para desenvolvimento da PE.

3 RISCOS E BENEFICIOS

3.1 Benefícios

O estudo tem como benefício principal a possibilidade de desenvolver um modelo preditivo de pré-eclâmpsia utilizando dados nacionais, contribuindo para a identificação precoce de gestantes em risco e auxiliando na melhoria dos desfechos materno-fetais, como também trará benefícios acadêmicos e científicos, com publicações em revistas nacionais e internacionais.

3.2 Riscos

Não existem riscos clínicos ou físicos aos participantes, uma vez que não haverá intervenção direta, apenas o uso de dados já existentes em prontuários eletrônicos. O risco identificado é de natureza informacional, referente à coleta, guarda e tratamento de dados clínicos em meio digital, especialmente considerando a utilização de armazenamento em dispositivos e em nuvem.

3.3 Prevenção e minimização dos riscos

- Os dados utilizados não conterão informações pessoais diretas (como nome, CPF ou endereço), apenas variáveis clínicas armazenadas em momento anterior a pesquisa.
- Cada participante será identificado apenas por um ID gerado no banco de dados no momento do cadastro do paciente na primeira consulta, garantindo o anonimato.
- Durante a coleta e análise, os dados estarão disponíveis em nuvem com criptografia ativada e acesso controlado por IP, restrito apenas ao pesquisador responsável.
- Após a conclusão da coleta, os arquivos serão transferidos para um dispositivo local protegido por senha e criptografia, e o acesso externo será bloqueado novamente pela clínica responsável.
- Os dados serão mantidos pelo período de 5 (cinco) anos após a finalização do estudo, conforme boas práticas científicas, sendo em seguida excluídos de forma definitiva.

Assim, os riscos relacionados ao uso em meio virtual são considerados mínimos e estão devidamente controlados, preservando a confidencialidade e garantindo a integridade das informações.

4 FUNDAMENTAÇÃO TEÓRICA

Para contextualizar e fundamentar o trabalho desenvolvido, esse capítulo aborda o conhecimento teórico sobre as teorias que fundamentam a abordagem do tema central do estudo."

4.1 Pré-Natal

O acompanhamento pré-natal configura-se como um pilar essencial para a redução da morbimortalidade materna e neonatal, integrando ações preventivas, diagnósticas e educativas. Estudos demonstram que consultas regulares diminuem em até 98% os riscos de óbitos por causas evitáveis, como hemorragias e infecções puerperais, mediante a identificação precoce de condições como pré-eclâmpsia e diabetes gestacional(Silva, I. E. C. *et al.*, 2025).

A atuação multidisciplinar no pré-natal revela-se crítica, com destaque para enfermeiros na detecção de riscos em Unidades Básicas de Saúde (UBS), onde consultas de enfermagem contribuem significativamente para reduzir complicações como partos prematuros, mediante monitoramento de pressão arterial, glicemia e orientações sobre aleitamento materno (SOUSA *et al.*, 2024). Em casos de diabetes gestacional, o acompanhamento especializado permite ajustes dietéticos e farmacológicos que diminuem intervenções cirúrgicas emergenciais(Oliveira; Silva, 2025).

Apesar dos avanços, desafios persistem na universalização do pré-natal qualificado. Pesquisas apontam que apenas 35% das gestantes em regiões periféricas completam as seis consultas mínimas recomendadas, muitas vezes devido a barreiras geográficas e culturais(QUEIROZ; SOUSA, 2024). Contudo, a carência de protocolos padronizados para abordagens específicas, como saúde bucal ou manejo de comorbidades, ainda limita a efetividade em escala nacional. Investimentos em capacitação profissional e tecnologias de monitoramento remoto surgem como alternativas promissoras, capazes de ampliar o acesso sem comprometer a qualidade do vínculo terapêutico(LIMA *et al.*, 2024).

4.2 Pré-eclâmpsia

A Pré-eclâmpsia (PE) é uma doença multissistêmica e uma das principais causas de morbidade e mortalidade materna e neonatal, contribuindo para cerca de 60.000 mortes maternas e mais de 500.000 nascimentos prematuros em todo o mundo a cada ano, sendo a condição marcada por hipertensão, podendo levar a várias outras complicações, incluindo convulsões (eclâmpsia), lesão renal aguda, disfunção hepática e até morte(SILVA, G. D. C. *et al.*, 2024).

Também pode resultar em desfechos neonatais adversos, como restrição do crescimento fetal e descolamento prematuro da placenta. Certos fatores aumentam o risco da gestante de desenvolver, como histórico de pré-eclâmpsia em gestações anteriores, obesidade, diabetes, hipertensão e gestações múltiplas, em que chega a atingir taxas variando de 8-20% para gêmeos e 12-34% para trigêmeos. Além disso, a condição é mais prevalente em populações com menor desenvolvimento, onde síndromes hipertensivas e complicações relacionadas são mais comuns (Ma'ayeh; Costantine, 2020).

Estudos demonstram que a administração de aspirina na dose de 150 mg por dia, de 11 a 14 semanas até 36 semanas da gestação, resultou em uma notável redução de 62% na incidência de PE prematura em gestantes com risco aumentado de desenvolver PE(Rolnik *et al.*, 2017). Como também a suplementação com cálcio demonstra eficácia na diminuição do risco de PE em 49% e hipertensão gestacional em 30%(Jaiswal *et al.*, 2024). Tais achados reforçam a necessidade de modelos preditivos que selezionem grávidas com risco aumentado de desenvolver PE ainda no início da gestação, momento em que a profilaxia se mostrou eficaz.

4.3 Biomarcadores

Os biomarcadores clínicos são ferramentas que ampliam a acurácia diagnóstica e fundamentam a medicina de precisão, como observado no câncer de pulmão, no qual assinaturas genéticas orientam terapias direcionadas(Xavier *et al.*, 2022).

Auxiliam no diagnóstico e monitoramento de doenças, permitindo que os médicos identifiquem e tratem condições de saúde de forma mais precisa e personalizada. No entanto, no Brasil, esses testes possuem um valor elevado, o que limita o acesso a muitos pacientes. Isso ocorre devido a uma combinação de fatores, incluindo a necessidade de importação, a complexidade tecnológica envolvida na sua detecção e análise, e os custos associados à regulamentação. Além disso, a infraestrutura de saúde no país pode não estar sempre equipada para realizar esses testes de forma eficiente, o que aumenta os custos operacionais e reflete no preço final para o paciente(Fonseca *et al.*, 2023)

4.4 Padrões arquitetônicos de softwares

4.5 API REST

Uma API REST (Representational State Transfer) é um padrão arquitetônico para comunicação entre sistemas que utiliza os princípios do protocolo HTTP para facilitar a comunicação entre clientes e servidores. Utiliza um conceito de “endpoints”, que são os pontos de comunicação do

sistema, no qual se pode enviar ou receber dados através de uma URL. Segundo (Gowda; Gowda, 2024), sua eficiência reside na adoção de práticas como: (1) design orientado a recursos, em que entidades (como usuários ou produtos) são acessadas via endpoints (ex: /api/v1/usuarios); (2) operações sem estado, garantindo que cada requisição contenha informações completas para processamento, eliminando dependências de sessões anteriores; e (3) uso explícito de métodos HTTP (GET, POST, PUT, DELETE) para mapear operações CRUD (Create, Read, Update, Delete). Essa abordagem traz benefícios significativos: escalabilidade, pois a ausência de conexões persistentes permite distribuir carga entre servidores; segurança robusta, com implementação de HTTPS e protocolos como JWT para autenticação; e flexibilidade, viabilizando integração facilitada entre sistemas. Além disso, técnicas como caching e paginação otimizam desempenho, reduzindo latência em cenários de alto tráfego(Gowda; Gowda, 2024). A adoção dessas práticas não apenas padroniza o desenvolvimento, mas também reduz riscos de vulnerabilidades e custos de manutenção, posicionando APIs REST como pilares essenciais para aplicações web modernas.

A integração de modelos preditivos por meio de APIs tem sido amplamente adotada para fornecer suporte à decisão médica em tempo real, facilitando a incorporação de inteligência artificial em sistemas hospitalares. APIs médicas permitem que dados clínicos sejam processados e analisados dinamicamente, proporcionando recomendações baseadas em evidências e aumentando a eficiência do diagnóstico(Mandel *et al.*, 2016). Além disso, a interoperabilidade entre sistemas eletrônicos de saúde possibilita que tais interoperabilidades sejam utilizadas em diversas plataformas, como prontuários eletrônicos e aplicativos móveis voltados para gestantes e profissionais de saúde(HIMSS – Healthcare Information and Management Systems Society, 2020). Portanto, unindo a previsão e a rápida integração facilitada com softwares já existentes, ou que podem vir a ser construídos, podemos tornar mais simples e efetivo esse acompanhamento clínico de gestantes em alto risco.

4.6 Machine Learning

Machine Learning (ML) é uma subárea da Inteligência artificial (IA) que utiliza algoritmos computacionais para identificar padrões em dados e realizar previsões, combinando métodos estatísticos e técnicas de ciência da computação. Na medicina, essa abordagem se divide em duas categorias principais: o aprendizado supervisionado, que prevê resultados (como o risco de infarto usando modelos como o Framingham Risk Score), e o aprendizado não supervisionado, que busca padrões ocultos em dados não rotulados (como a identificação de subtipos de doenças

para medicina personalizada). Apesar do potencial, desafios significativos limitam sua aplicação prática: a necessidade de grandes conjuntos de dados diversificados, a dificuldade em selecionar variáveis biologicamente relevantes e a validação rigorosa dos modelos para garantir precisão em diferentes populações(Deo, 2015).

Existem diversos modelos de ML, como: Random Forest, SVM (Support Vector Machine), Redes Neurais, entre outros. São algoritmos que aprendem com dados de entrada para fazer previsões ou decisões. Possui uma ampla gama de aplicações em vários campos, incluindo saúde, finanças, marketing e muito mais. Eles podem melhorar os produtos, otimizar processos e aprimorar as capacidades de tomada de decisão nas tarefas diárias.(Sarker, 2021).

O uso de ML tem se destacado na área da saúde como uma alternativa promissora para a predição de doenças, oferecendo maior precisão na análise de grandes volumes de dados clínicos(Shickel *et al.*, 2018). Modelos como redes neurais, árvores de decisão e regressão logística têm sido aplicados na predição de complicações gestacionais, demonstrando capacidade de identificar padrões complexos que não seriam facilmente detectáveis por métodos convencionais(Artzi *et al.*, 2020). Estudos recentes indicam que modelos baseados em aprendizado supervisionado conseguem atingir altas taxas de sensibilidade e especificidade, quando treinados com bases de dados clínicos robustas e diversificadas(Marić *et al.*, 2020).

5 ESTADO DA ARTE

Métodos tradicionais de predição para a PE incluem a análise de fatores de risco clínicos, como idade materna, índice de massa corporal (IMC), antecedentes obstétricos e medidas da pressão arterial(American College of Obstetricians and Gynecologists (ACOG), 2020). Além disso, exames laboratoriais e ultrassonográficos, como a dosagem do fator de crescimento placentário (PLGF) e a medida da artéria uterina por Doppler, têm sido utilizados para estimar o risco de desenvolvimento da condição(Poon; Nicolaides, 2014). No entanto, essas abordagens apresentam limitações, como alto custo e baixa acessibilidade, dificultando sua ampla adoção em países de baixa e média renda(World Health Organization, 2019).

As calculadoras de riscos também são utilizadas para prever a PE e empregam métodos de regressão integrando um conjunto de variáveis preditivas, incluindo características maternas, pressão arterial média (PAM), índice de pulsatilidade da artéria uterina (UTA-PI), PLGF e proteína plasmática A associada à gravidez (PAPP-A) para estimar riscos obstétricos. Desenvolvidas em colaboração entre a Fetal Medicine Foundation (FMF) e o Hospital Clínic de Barcelona, essas ferramentas demonstraram maior precisão diagnóstica em comparação com métodos tradicionais, além de serem a única abordagem com validação internacional(Riishede *et al.*, 2023).

O estudo de (Castro Rezende *et al.*, 2024), realizado no brasil, mostra que mesmo validados, métodos como esse não levam em conta a população local para calibrar seus parâmetros fundamentais, o que é essencial para uma boa precisão em tipos de populações distintas, no qual o resultado é uma taxa considerada alta de falsos positivos (FPR) de 24,4%. Quando comparamos com outros modelos preditivos que utilizam ML, que obtêm resultados de FPR abaixo de 15%(Gil *et al.*, 2024). Com isso podemos observar a discrepância desses resultados, demonstrando que a falta de dados personalizados para a população afeta expressivamente o nível de precisão do resultado. Apesar disso, o modelo da FMF mantém relevância clínica por sua validação internacional e simplicidade operacional, sendo amplamente adotado em contextos de recursos limitados e frequentemente referenciadas por outros autores que desenvolvem modelos de aprendizado de máquina como um padrão para validação de sua acurácia, por ser um padrão aceito e validado(Torres-Torres *et al.*, 2024).

Demonstrando o potencial das tecnologias de IA na transformação do cuidado pré-natal um estudo desenvolveu usando Python, scikit-learn e TensorFlow, com dados analisados por meio

do SPSS 22 uma versão do IBM SPSS Statistics para prever o risco de PE. Enquanto algoritmos de deep learning e Extra Trees Classifier, foram empregados para avaliar o poder preditivo de diferentes variáveis. O estudo utilizou técnicas como aumento de gradiente e otimização de hiper parâmetros para melhorar o desempenho do modelo, que demonstrou uma sensibilidade de 73,7% e especificidade de 92,7%, indicando sua eficácia na distinção entre casos positivos e negativos para pré-eclâmpsia(Bülez *et al.*, 2024).

Recente estudo chinês testou cinco algoritmos de classificação diferentes, Logistic Regression, Extra Trees Classifier, Voting Classifier, Gaussian Process Classifier e Stacking Classifier. O Stacking Classifier apresentou o melhor desempenho na previsão da pré-eclâmpsia prematura, com uma área sob a curva (AUC) de 0,884. no qual incorpora várias características maternas, como idade, altura, peso antes da gravidez e biomarcadores clínicos, como PAM, UTA-PI, PAPP-A e PLGF, utilizando da ferramenta SHAP (ShaPley Additive Explanations) para explicar as previsões do modelo e oferecer mais transparência do modelo, o que torna mais seguro a aplicação em contextos clínicos, porém ressalta a necessidade de validação desse método por considerar dados de apenas um único centro local(Li *et al.*, 2024).

Portanto, fica evidente a viabilidade e relevância do desenvolvimento de uma nova ferramenta preditiva para avaliação do risco de pré-eclâmpsia baseada em algoritmos de machine learning, utilizando dados clínicos específicos da nossa população. Essa abordagem personalizada tem o potencial de superar as limitações dos modelos atuais, especialmente no que se refere à taxa de falsos positivos e à acurácia em contextos locais, contribuindo significativamente para a tomada de decisão clínica mais precisa e eficaz e facilitar a implementação dessas ferramentas em diferentes realidades socioeconômicas. Diante deste cenário, construímos um modelo de ML, o Lightgbm, implementado em uma API REST para prever a PE, elegemos esse modelo após uma revisão sistemática da eficácia do machine learning na predição da PE no geral, e para a nossa realidade do tamanho da amostra que iremos trabalhar os algoritmos da família do gradient boosting(GBM) se mostraram sempre acertivos e sem complexidade para o treinamento.

6 PPML – API REST FULL

Foi desenvolvida em Python 3.9+ utilizando FastAPI como framework web assíncrono, com validação automática de entrada via Pydantic e documentação OpenAPI/Swagger nativa. A comunicação ocorre via HTTP/HTTPS com payloads JSON, suportando deploy containerizado utilizando Uvicorn.

O modelo final implementado foi o LightGBM(LGBMClassifier) de Gradiente Boosting otimizado para classificação binária, por demonstrar resultados com mais precisão com o tamanho da amostra disponível, treinado com 24 features clínicas/obstétricas(peso, IMC, diabetes, hipertensão, percentis Doppler, comorbidades). Configurando com n_estimators=500, learning_rate=0.1, num_leaves=31, max_depth=6 e regularização (reg_alpha/lambda=0.1), alcançou AUC-ROC 0.9669, precisão/recall/F1 0.9091 no conjunto de teste independente(22+22 casos), com early stopping em 800 iterações.

O pré-processamento segue pipeline customizada com pandas (manipulação), scikit-learn StandardScaler (normalização), mapeamentos categóricos (raça: Branco=1/Pardo=2/Preto=3; histórico diabetes), extração de comorbidades via regex, preenchimento NaN (0/moda), clipping realista (peso 40-150kg) e garantia de 24 features fixas.

O treinamento incorporou SMOTE (balanceamento 15.9%→50% positivos), GroupShuffleSplit por paciente (evita data leakage), data augmentation gaussiana (+40% nas features contínuas) e splits estratificados (treino 72%, validação 13%, teste 15%). Avaliação incluiu ROC-AUC, matriz confusão, curvas aprendizado e SHAP-like feature importance (top: peso, diabetes, idadeGestacional).

Endpoints expõem predict_lgbm (probabilidade + classificação risco), health (status modelos) e features (esquema entrada), com threshold 0.5 para alto/baixo risco e tratamento robusto de exceções. disponibilizaremos o serviço através de um servidor Linux de alta disponibilidade, hospedado no Heroku, uma plataforma em nuvem de baixo custo, podendo também ser utilizado como ambiente de treinamento e teste dos modelos.

Quadro 1: Dados de entrada para predição

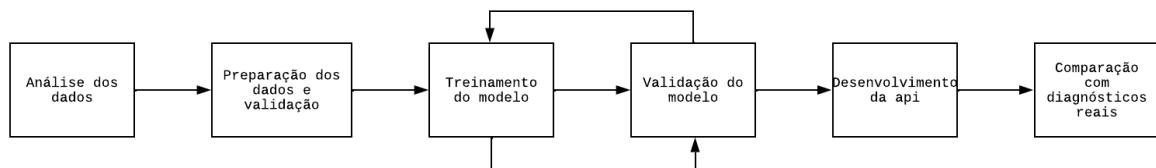
Peso (kg)
Índice de Massa Corporal (IMC)
Diabetes
Hipertensão
Peso no primeiro trimestre
Origem racial
Histórico familiar de diabetes
Índice de pulsatilidade médio da artéria uterina (IP médio)
História obstétrica anterior
Perdas gestacionais
Tabagismo durante a gravidez
Idade gestacional
Idade gestacional corrigida
Peso fetal
Percentil da artéria uterina
Percentil do peso fetal
Hipertensão pré-existente
Endometriose
Hipotireoidismo
Doença inflamatória intestinal (IIC)
Trombofilia
Retocolite
Gastrite
Outras doenças associadas

Fonte: Adaptado de Fetal Medicine Foundation, 2025.

O projeto foi organizado em pastas funcionais básicas para facilitar manutenção e desenvolvimento ágil, no qual a pasta API contém os endpoints RESTful, gerenciando recebimento e validação dos dados e respostas. já a pasta DATA cuida do pré-processamento, limpeza e transformação dos dados clínicos necessários para alimentar o modelo. A pasta MODELS armazena o modelo treinado, o LightGBM serializado, e scripts para carregamento e inferência. a pasta NOTEBOOKS é usada para experimentação, treinamento, avaliação e geração dos modelos, incluindo análise exploratória e métricas. A organização do código foi pensada para seguir uma organização simples, seguindo os princípios do Clean code, garantindo separação das responsabilidades e modularidade.

7 METODOLOGIA

A metodologia do desenvolvimento da PPML seguirá o processo de pipeline de Machine Learning, baseado no modelo proposto por (Kreuzberger; Kühl; Hirschl, 2023), adaptado à nossa proposta de aplicação clínica.



7.1 Análise dos dados

Os dados para treinamento do modelo serão baseados em dados clínicos e exames ultrassonográficos armazenados no banco de dados da clínica privada EMBRION que nos autorizou o seu uso através de uma carta de consentimento (anexo B).

REFERÊNCIAS

- ALCINO, M. S. *et al.* Inteligência artificial e saúde materna: a experiência da Caren em Goiás. **Journal of Health Informatics**, v. 16, Especial, 2024. DOI: [10.59681/2175-4411.v16.iEspecial.2024.1271](https://doi.org/10.59681/2175-4411.v16.iEspecial.2024.1271).
- AMERICAN COLLEGE OF OBSTETRICIANS AND GYNECOLOGISTS (ACOG). **Hypertension in Pregnancy**. [S. l.: s. n.], 2020. <https://www.acog.org/>. Acesso em: 13 fev. 2025.
- ARTZI, Natalie S. *et al.* Prediction of gestational diabetes based on nationwide electronic health records. **Nature Medicine**, v. 26, n. 1, p. 71–76, jan. 2020. DOI: [10.1038/s41591-019-0724-8](https://doi.org/10.1038/s41591-019-0724-8).
- BÜLEZ, A. *et al.* Artificial Intelligence in Early Diagnosis of Preeclampsia. **Nigerian Journal of Clinical Practice**, v. 27, n. 3, p. 383, mar. 2024.
- CASTRO REZENDE, Karina Bilda de *et al.* Performance of the first-trimester Fetal Medicine Foundation competing risks model for preeclampsia prediction: an external validation study in Brazil. **AJOG Global Reports**, v. 4, n. 2, p. 100346, 2024. DOI: [10.1016/j.xagr.2024.100346](https://doi.org/10.1016/j.xagr.2024.100346). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/38694483/>.
- CHAPPELL, Lucy C. *et al.* Pre-eclampsia. **The Lancet**, v. 398, n. 10297, p. 341–354, 2021. DOI: [10.1016/S0140-6736\(20\)32335-7](https://doi.org/10.1016/S0140-6736(20)32335-7).
- DEO, Rahul C. Machine Learning in Medicine. **Circulation**, v. 132, n. 20, p. 1920–1930, 2015. DOI: [10.1161/CIRCULATIONAHA.115.001593](https://doi.org/10.1161/CIRCULATIONAHA.115.001593).
- FONSECA, Evelyn *et al.* BIOMARCADORES NO DIAGNÓSTICO DE DOENÇAS. **REVISTA FOCO**, v. 16, e3813, dez. 2023. DOI: [10.54751/revistafoco.v16n12-014](https://doi.org/10.54751/revistafoco.v16n12-014).
- GAROVIC, Vesna D. *et al.* Hypertension in Pregnancy: Diagnosis, Blood Pressure Goals, and Pharmacotherapy: A Scientific Statement From the American Heart Association. **Hypertension**, v. 79, n. 2, e21–e41, 2022. DOI: [10.1161/HYP.000000000000208](https://doi.org/10.1161/HYP.000000000000208).
- GIL, M. M. *et al.* Validating a machine-learning model for first-trimester prediction of pre-eclampsia using the cohort from the PREVAL study. **Ultrasound in Obstetrics & Gynecology**, v. 63, n. 1, p. 68–74, 2024. DOI: [10.1002/uog.27478](https://doi.org/10.1002/uog.27478). Disponível em: <https://obgyn.onlinelibrary.wiley.com/doi/10.1002/uog.27478>.

GOWDA, Priyanka; GOWDA, Ashwath Narayana. Best Practices in REST API Design for Enhanced Scalability and Security, 2024. Disponível em:
https://www.researchgate.net/profile/Priyanka-Gowda-Ashwath-Narayana-Gowda-2/publication/383599280_Best_Practices_in_REST_API_Design_for_Enhanced_Scalability_and_Security/links/67a54292461fb56424cc95a0/Best-Practices-in-REST-API-Design-for-Enhanced-Scalability-and-Security.pdf.

HIMSS – HEALTHCARE INFORMATION AND MANAGEMENT SYSTEMS SOCIETY. **The role of APIs in modern healthcare.** [S. l.: s. n.], 2020. HIMSS Report. Acesso em: 13 fev. 2025. Disponível em: <https://www.himss.org/>.

JAISWAL, Vikash *et al.* Association between calcium supplementation and gestational hypertension, and preeclampsia: A Meta-analysis of 26 randomized controlled trials. **Current Problems in Cardiology**, v. 49, n. 3, p. 102217, 2024. ISSN 0146-2806. DOI: [10.1016/j.cpcardiol.2023.102217](https://doi.org/10.1016/j.cpcardiol.2023.102217). Disponível em:
<https://www.sciencedirect.com/science/article/pii/S0146280623006345>.

KREUZBERGER, Daniel; KÜHL, Niklas; HIRSCHL, Stefan. Machine Learning Operations (MLOps): Overview, Definition, and Architecture. **IEEE Access**, v. 11, p. 1–1, 2023. DOI: [10.1109/ACCESS.2023.3241234](https://doi.org/10.1109/ACCESS.2023.3241234).

LI, Taishun *et al.* Prediction model of preeclampsia using machine learning based methods: a population based cohort study in China. **Frontiers in Endocrinology**, v. 15, p. 1345573, jun. 2024. DOI: [10.3389/fendo.2024.1345573](https://doi.org/10.3389/fendo.2024.1345573). Disponível em: <https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2024.1345573/full>.

LIMA, Mirella Maria de *et al.* A importância do acompanhamento do pré-natal na Atenção Básica. **Brazilian Journal of Implantology and Health Sciences**, v. 6, n. 4, p. 2457–2468, abr. 2024. DOI: [10.36557/2674-8169.2024v6n4p2457-2468](https://doi.org/10.36557/2674-8169.2024v6n4p2457-2468). Disponível em:
<https://bjih.s.emnuvens.com.br/bjih/article/view/2000>.

MA'AYEH, Marwan; COSTANTINE, Mohamed M. Prevention of preeclampsia. **Seminars in Fetal and Neonatal Medicine**, v. 25, n. 5, p. 101123, out. 2020. Epub 2020 Jun 2. DOI: [10.1016/j.siny.2020.101123](https://doi.org/10.1016/j.siny.2020.101123).

MANDEL, Joshua C. *et al.* SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. **Journal of the American Medical Informatics Association**, v. 23, n. 5, p. 899–908, set. 2016. Epub 2016 Feb 17. DOI: [10.1093/jamia/ocv189](https://doi.org/10.1093/jamia/ocv189).

MARIĆ, Ivan *et al.* Early prediction of preeclampsia via machine learning. **American Journal of Obstetrics & Gynecology MFM**, v. 2, n. 2, p. 100100, maio 2020. DOI: [10.1016/j.ajogmf.2020.100100](https://doi.org/10.1016/j.ajogmf.2020.100100).

OLIVEIRA, Fabricia Araújo; SILVA, Gustavo Iltemberg Sousa. Intervenções de enfermagem no pré-natal de alto risco: revisão integrativa. **Revista Saber Digital**, v. 18, n. 2, e20251812, jun. 2025. DOI: [10.24859/SaberDigital.2025v18n2.1745](https://doi.org/10.24859/SaberDigital.2025v18n2.1745).

POON, L. C.; NICOLAIDES, K. H. Early Prediction of Preeclampsia. **Obstetrics and Gynecology International**, v. 2014, p. 1–11, 2014. DOI: [10.1155/2014/297397](https://doi.org/10.1155/2014/297397).

QUEIROZ, Yvine; SOUSA, Milena. Empoderamento materno: a importância de abordar sobre via de parto durante pré-natal. **Revista Brasileira de Filosofia e História**, v. 13, p. 2322–2330, fev. 2024. DOI: [10.18378/rbfh.v13i1.10360](https://doi.org/10.18378/rbfh.v13i1.10360).

REZENDE, K. B. C. *et al.* Performance of the first-trimester Fetal Medicine Foundation competing risks model for preeclampsia prediction: an external validation study in Brazil. **AJOG Global Reports**, v. 4, n. 2, p. 100346, 2024. DOI: [10.1016/j.xagr.2024.100346](https://doi.org/10.1016/j.xagr.2024.100346).

RIISHEDE, I. *et al.* Pre-eclampsia screening in Denmark (PRESIDE): national validation study. **Ultrasound in Obstetrics & Gynecology**, v. 61, n. 6, p. 682–690, maio 2023. DOI: [10.1002/uog.26183](https://doi.org/10.1002/uog.26183). Disponível em: <https://obgyn.onlinelibrary.wiley.com/doi/10.1002/uog.26183>.

ROLNIK, Daniel L. *et al.* ASPRE trial: performance of screening for preterm pre-eclampsia. **Ultrasound in Obstetrics & Gynecology**, v. 50, n. 4, p. 492–495, ago. 2017. DOI: [10.1002/uog.18816](https://doi.org/10.1002/uog.18816).

SARKER, Iqbal H. Machine Learning: Algorithms, Real-World Applications and Research Directions. **SN Computer Science**, Springer, v. 2, n. 3, p. 160, 2021. ISSN 2661-8907. DOI: [10.1007/s42979-021-00592-x](https://doi.org/10.1007/s42979-021-00592-x).

SHICKEL, Benjamin *et al.* Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. **IEEE Journal of Biomedical and Health Informatics**, v. 22, n. 5, p. 1589–1604, set. 2018. DOI: [10.1109/JBHI.2017.2767063](https://doi.org/10.1109/JBHI.2017.2767063).

SILVA, Guilherme Dias Coelho *et al.* Impactos da pré-eclâmpsia na gravidez. **Cuadernos de Educación y Desarrollo**, v. 16, n. 2, 2024. Acesso em: 2025-11-24. Disponível em: <https://ojs.cuadernoseducacion.com/ojs/index.php/ced/article/view/2659/2098>.

SILVA, Idson Emanuel Cavalcanti *et al.* PERFIL DA MORTALIDADE MATERNA: O PAPEL CRÍTICO DA ASSISTÊNCIA PRÉ-NATAL NA REDUÇÃO DE ÓBITOS MATERNOS NO BRASIL. **Arquivos de Ciências da Saúde da UNIPAR**, v. 28, n. 3, p. 951–967, 2025. DOI: [10.25110/arqsaude.v28i3.2024-10994](https://doi.org/10.25110/arqsaude.v28i3.2024-10994). Disponível em: <https://www.revistas.unipar.br/index.php/saude/article/view/10994>.

SILVA, Mayara Amanda da. **Avaliação de Modelos de Aprendizado de Máquina para Classificação de Gestantes e Predição de Gravidez de Risco Usando o Histórico de Consultas Médicas**. 2021. Diss. (Mestrado) – Universidade Federal de Juiz de Fora, Juiz de Fora, MG. Orientador: Alex Borges Vieira; Coorientadores: Artur Ziviani, Heder Soares Bernardino. Disponível em: <https://repositorio.ufjf.br/jspui/bitstream/ufjf/13763/1/mayaraamandadasilva.pdf>.

TORRES-TORRES, J. *et al.* Performance of machine-learning approach for prediction of pre-eclampsia in a middle-income country. **Ultrasound in Obstetrics & Gynecology**, v. 63, n. 3, p. 350–357, 2024. DOI: <https://doi.org/10.1002/uog.27510>. eprint: <https://obgyn.onlinelibrary.wiley.com/doi/pdf/10.1002/uog.27510>. Disponível em: <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/uog.27510>.

VIELLAS, Elaine *et al.* Prenatal care in Brazil. **Cadernos de Saúde Pública**, v. 30, s85–s100, ago. 2014. DOI: [10.1590/0102-311X00126013](https://doi.org/10.1590/0102-311X00126013).

WORLD HEALTH ORGANIZATION. **WHO recommendations for prevention and treatment of pre-eclampsia and eclampsia**. Geneva: World Health Organization, 2019. <https://www.who.int/publications/i/item/9789241548335>. Accessed 13 Feb 2025.

XAVIER, R. F. *et al.* Perspectives in the treatment of lung cancer: analysis of signaling pathways and biomarkers. **Research, Society and Development**, v. 11, n. 7, e5411722903, 2022. DOI: [10.33448/rsd-v11i7.22903](https://rsdjournal.org/index.php/rsd/article/view/22903). Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/22903>. Acesso em: 9 mar. 2025.

APÊNDICE A – TÍTULO DO APÊNDICE

Conteúdo do Apêndice A...

Conteúdo do Apêndice A...