

# Marcos Benício de Andrade Alonso

Brasileiro, 30 anos – Niterói- RJ

**Telefone:** (21) 9 9564-7583

**Portifólio/Blog:** <https://marcosbenicio.github.io/>

**Linkdin:** <https://www.linkedin.com/in/marcos-benicio-de-andrade-alonso/>

**GitHub:** <https://github.com/marcosbenicio>

**E-mail:** marcosbenicio@id.uff.br

## Objetivo Profissional

---

Durante meus anos na academia, busquei unir minha paixão pela física e pela computação, participando de projetos na interseção de ambas as ciências. Atualmente, busco a oportunidade de utilizar meus conhecimentos de física, estatística e machine learning em um ambiente dinâmico de colaboração e inovação para solucionar problemas de forma data-driven.

## Formação Acadêmica

---

### 1. Doutorado em Física (Universidade Federal do Rio de Janeiro, 2021 – Atual)

Pesquisa na área de redes neurais, com ênfase em redes neurais bayesianas, para a solução de equações diferenciais parciais usadas na modelagem de problemas de dinâmica dos fluidos.

### 2. Mestrado em Física (Universidade Federal Fluminense, 2019-2021)

Pesquisa feita na área de Complexidade Computacional e Mecânica Quântica, onde analisei algoritmos quânticos e seus recursos necessários para rodá-los num computador quântico. Durante esse período publiquei um artigo sobre esse tema no Quantum Journal: [artigo](#).

### 3. Bacharel em Física (Universidade Federal Fluminense, 2014-2018)

Particpei de pesquisa na área de epidemiologia, desenvolvendo modelos matemáticos e simulações computacionais para investigar condições numa população para o espalhamento de epidemias.

## Projetos

---

Os projetos a seguir foram desenvolvidos durante cursos que fiz sobre ciência de dados e durante meus anos na universidade. A função desses projetos é demonstrar meu domínio em programação e algumas ferramentas essenciais para um cientista de dados, além do meu entendimento sobre machine learning e redes neurais.

- **[Modelos De Machine Learning](#)**

Notas sobre a teoria e algoritmos em [Python](#) para os modelos de [Regressão Linear](#), [Regressão Logística](#), [Arvores de Decisão](#), [Random Forest](#), [XGBoost](#) e [Redes Convolucionais](#). Além disso, escrevi notas sobre métricas usadas em modelos de classificação e regressão, e sobre o uso de [Docker](#), [Pickle](#) e [Flask](#) para o deploy desses modelos de machine learning. Dois projetos de Data Science foram desenvolvidos usando esses conceitos, sendo eles descritos a seguir:

- **Regressão - Duração de Viagem de Taxi em Nova York**

Projeto de Regressão para prever o tempo de viagem de táxi em Nova York utilizando grande volume de dados. Foi realizada uma extensa análise exploratória, e novas features foram criadas utilizando o dataset de previsão do tempo, empregando o algoritmo k-means para a criação de clusters para os pontos de embarque e desembarque, horários de congestionamento e informações geoespaciais, como o cálculo de geodésicas através da biblioteca geopandas. Ao final, foram treinados os modelos XGBoost e Árvore de Decisão para regressão. O modelo com a melhor métrica Root Mean Squared Logarithmic Error (RMSLE) foi usado para criar uma imagem Docker com Flask, a fim de enviar requisições HTTP para o deployment utilizando a plataforma Render.

- **Classificação – Fatores Para Identificar Diabetes**

Projeto para identificar os fatores mais relevantes para a predição de diabetes em pacientes. Foi realizada uma análise exploratória da base de dados para a seleção de features com maior relevância para o treinamento de modelos de machine learning. Foram utilizados os modelos de Regressão Logística, Árvores de Decisão e Random Forest. Para comparar o desempenho dos modelos, foram empregadas as métricas F1 score, Precision, Recall, Matriz de Confusão, AUC. O desafio dessa base de dados foi reduzir a taxa de falsos negativos (indivíduos classificados incorretamente como não tendo diabetes). Após o treinamento do modelo, foi criada uma imagem Docker com Flask para enviar requisições HTTP para o deployment utilizando a plataforma Render.

- **Engenharia de Dados**

Notas sobre o uso de Docker, PostgreSQL, ferramentas de orquestração de pipeline ETL, como Airflow e Mage, o uso da Google Cloud Platform, a criação de um pipeline ELT utilizando dbt, e como usar PySpark para lidar com grandes volumes de dados. No momento, estou desenvolvendo um projeto de engenharia de dados com essas ferramentas para colocar em prática meus estudos.

- **Physics Informed Neural Networks**

Projeto onde reproduzi alguns dos resultados propostos neste artigo, fazendo uso de redes neurais para solucionar equações diferenciais parciais de famosos problemas físicos, como os envolvendo as equações de Schrödinger, Burger e Stokes. Foram usadas as bibliotecas Tensorflow e Keras em Python para esse projeto, além de boas práticas de OOP (Programação Orientada a Objetos).

- **Modelos Epidemiológicos**

Projeto desenvolvido na iniciação científica durante a graduação. Foram utilizados diferentes modelos epidemiológicos para compreender a evolução de uma epidemia para uma pandemia. Além disso, um novo modelo foi desenvolvido levando em consideração a adição de uma vacina na população. Os códigos foram feitos em C, os gráficos gerados no gnuplot (ferramenta CLI para Linux), e as simulações rodadas no cluster da Universidade. Para acessar remotamente o cluster, utilizando o protocolo SSH, foi necessário conhecimento sobre o sistema operacional Linux.

- **Blog**

Blog criado para escrita de artigos e projetos de Data Science. O blog foi feito em HTML, CSS, Javascript e jQuery.

## **Experiência Profissional**

---

- **Estágio Docência na disciplina Informação e Computação Quântica, UFF (2019 – 2020)**  
Monitor no curso Informação e Computação Quântica no mestrado de Física.
- **Pesquisador Bolsista no Ensino Superior, UFF-UFRJ (2019 – Atual)**  
Atuação como pesquisador bolsista CNPq na UFF durante o mestrado e CAPES na UFRJ durante o doutorado.
- **Iniciação Científica, UFF (2016 – 2018)**  
Atuação como aluno pesquisador CNPq junto ao grupo de pesquisa de Redes Complexas da UFF.
- **Professor de Física no Ensino Médio, Colégio e Curso Migel Couto (2016 - 2018)**

## Competências e Habilidades

---

- **Python:** Pandas, Sklearn, Matplotlib, Seaborn, Numpy, Scipy, Pyenv, Venv, Tensorflow, Keras;
- **C;**
- **SQL:** PostgreSQL, duckdb;
- **Google Cloud Platform:** Virtual Machine Instances, Cloud Storage, BigQuery;
- **Versionamento:** Git, GitHub;
- **Ferramentas:** Docker, Linux, Shell Scripting, Airflow;
- **Conhecimentos Teóricos:** Estatística, Cálculos, Álgebra Linear, modelos de Machine Learning e Redes Neurais ( Perceptron Multi-Camadas, Bayesianas, Convolutional)
- **Inglês Fluente;**