

Franklin Ferreira da Silva Filho

Brasileiro, 29 anos - Vitória de Santo Antão - PE

Telefone: (81) 99703-3380

Portfólio: <https://franklinzppa.github.io/>

Linkedin: <https://www.linkedin.com/in/franklin-ferreira-filho>

GitHub: <https://github.com/franklinzppa>

E-mail: franklin.ferreira@ufpe.br

OBJETIVO PROFISSIONAL

Atuar como cientista de dados, utilizando estatística e *machine learning* para resolver problemas de negócio de forma *data driven*. Desenvolver minha carreira em um ambiente dinâmico de colaboração e inovação.

FORMAÇÃO ACADÊMICA

- **Doutorando em Física** (Universidade Federal de Pernambuco, 10/2020 - Atual)
Desenvolvendo pesquisa na área de simulação computacional de materiais com potenciais treinados por machine learning aplicado a transporte térmico e eletrônico.
- **Mestrado em Física** (Universidade Federal Rural de Pernambuco, 09/2018 - 07/2020)
Pesquisa na área de Mecânica Quântica, estudando confinamento em nanoestruturas 1D.
- **Intercâmbio** (University of Manchester (UK), 09/2014 - 09/2015)
Período de intercâmbio na *University of Manchester* pelo Ciências sem Fronteiras.
- **Licenciatura em Química** (Universidade Federal Rural de Pernambuco, 03/2012 - 08/2018)
Pesquisa na área de Mecânica Quântica, estudando o formalismo de confinamento em variedades.

PROJETOS DE CIÊNCIAS DE DADOS

Os projetos aqui citados foram desenvolvidos em etapas do ciclo CRISP-DM: entendimento do negócio, pré-processamento, análise exploratória, tratamento dos dados, treinamento de modelos, otimização de hiperparâmetros, avaliação e *deploy*.

- **Classificação de Diabetes**
Projeto de classificação de pacientes diabéticos em função de marcadores de saúde. Foi criado um modelo heurístico para comparações com o modelo de aprendizado de máquina, e estruturado um problema de negócio que simulava o impacto financeiro do modelo preditivo. As transformações realizadas foram automatizadas em uma pipeline. A otimização de hiperparâmetros dos modelos treinados (RandomForest, SVM, etc) foi realizada com validação cruzada (RandomSearchCV). A avaliação do modelo incluiu a análise da matriz de confusão, ROC, curva lift e o deploy foi realizado no HugginFace.
- **Clusterização de Clientes**
Projeto de segmentação de clientes. Foi utilizado o algoritmo de *K-prototypes* para o modelar tanto variáveis numéricas quanto categóricas. Realizou-se uma análise dos

clusters obtidos e detectado possíveis estratégias de negócio. Um modelo alternativo foi construído com K-means em conjunto com PCA para análise visual dos clusters obtidos.

- **Regressão de Renda**

Projeto de previsão de renda utilizando dados do Censo 2000. Projeto com grande volume de dados extraídos com SQL (BigQuery). Uma possível aplicação desse projeto é na verificação de informações autodeclaradas, a fim de evitar fraudes. No tratamento, foi importante o uso de transformações (*log transform*) e tratamento de *outliers*. Também foi utilizado ferramentas de análise geoespacial (geopandas) na análise dos dados.

- **Classificação de SPAM**

Projeto de detecção de SPAM utilizando técnicas de NLP. Os dados foram obtidos via ETL e foram realizadas etapas de limpeza dos dados e exploração (análise de sentimentos e wordcloud). Além dos modelos usuais (LogisticRegression, SVM, etc) foi treinado um modelo de redes neurais (MLP) utilizando o Keras.

- **Modelos de Machine Learning**

Projeto de reescrita de algoritmos de machine learning para entendimento teórico. Entre os modelos descritos estão KNN, regressão linear, árvores de decisão, K-Means e Logistic Regression. Foram incluídos vídeos de explicação dos modelos (no momento, [KNN](#)).

- **Blog de Data Science**

Escrita de artigos de divulgação científica de ciência de dados e estatística. Até o momento inclui a postagem sobre intervalos de confiança e sobre probabilidade.

EXPERIÊNCIA PROFISSIONAL

- **Professor de Física**, Colégio Diogo de Braga (02/2022 – 06/2022);
Atuação como professor substituto de física do ensino médio.
- **Monitoria de Química Inorgânica LI**, UFRPE (02/2016 - 02/2017);
Monitor em aulas teóricas e experimentais.
- **Estágio docência na disciplina de Física para Biólogos**, UFRPE (07/2019 - 12/2019);
Professor em disciplina do curso de Bacharelado em Biologia.
- **Pesquisador**, UFRPE (03/2012 - 08/2018);
Atuação como pesquisador PIBIC no LAQIS-UFRPE construindo biossensores e como químico teórico-computacional no LNP-UFRPE.

COMPETÊNCIAS E HABILIDADES

- **Python**: Pandas, Sklearn, Matplotlib, Seaborn, Numpy, Scipy, OOP, Pyenv, Venv;
- **SQL**: BigQuery, SQLite;
- **Versionamento**: Git, GitHub;
- **Cloud AWS**: Sagemaker, Athena, S3, Glue, Lambda, ECS, IAM;
- Outras ferramentas: Linux, Shell Scripting, HTML, CSS, Flask, C;
- Inglês fluente.