

Clébio Marques de Oliveira Júnior

Email: clebiomojunior@gmail.com

LinkedIn: <https://www.linkedin.com/in/clebiojunior>

GitHub: <https://www.github.com/juniorcl>

Resumo Profissional

Cientista de Dados com mais de 4 anos de experiência em machine learning, processamento de linguagem natural (NLP) e análise de dados. Atuação em projetos de credit scoring, previsão de vendas, segmentação de clientes, previsão de churn e modelagem comportamental. Experiência prática com Python, pandas, scikit-learn, spaCy, OCR, regex, FAISS e pipelines de dados de ponta a ponta.

Experiência Profissional

Cientista de Dados – Vert Analytics

Outubro 2024 – Presente

- Desenvolveu sistema de recomendação de defesas jurídicas utilizando técnicas de NLP, embeddings e FAISS.
- Implementou pipelines de extração de dados a partir de documentos PDF utilizando OCR (Tesseract) e expressões regulares.
- Construiu pipelines de processamento de texto para comentários em redes sociais utilizando spaCy, incluindo tokenização, lematização, normalização e correção gramatical.

Cientista de Dados – Datarisk

Janeiro 2022 – Agosto 2024

- Desenvolveu modelo de credit scoring para apoio à decisão de aprovação de crédito, melhorando a avaliação de risco e reduzindo inadimplência.
- Criou modelo de previsão de vendas com regressão, reduzindo o erro médio absoluto (MAE) em 12%.
- Realizou segmentação de clientes utilizando algoritmos de clustering para apoiar campanhas de marketing direcionadas.
- Implementou modelos preditivos comportamentais para identificar propensão a churn, risco de crédito e instabilidade laboral.

Cientista de Dados – Be.X!

Março 2021 – Janeiro 2022

- Extraiu e estruturou dados textuais não estruturados utilizando expressões regulares e técnicas de processamento de linguagem natural.
 - Projetou modelos de detecção de anomalias baseados em regras alinhadas a requisitos de negócio.
 - Desenvolveu modelos de previsão de atrasos em entregas utilizando métricas baseadas em ranking.
-

Formação Acadêmica

Mestrado em Ciências Naturais (Ciências Ambientais)

Universidade Estadual do Norte Fluminense (UENF)

2017 – 2019

Licenciatura em Física

Instituto Federal Fluminense

2010 – 2016

Certificações

- Estatística para Ciência de Dados e Machine Learning
 - Sumarização de Textos com Processamento de Linguagem Natural
 - Clustering: Extrair Padrões de Dados
 - Reconhecimento de Textos com OCR e Python
 - Machine Learning e Data Science com Python do Zero ao Avançado
-

Habilidades Técnicas

Linguagens de Programação: Python, SQL

Bibliotecas e Frameworks: pandas, scikit-learn, spaCy, FAISS

Machine Learning: regressão, classificação, clustering, análise de séries temporais

Processamento de Linguagem Natural: tokenização, lematização, análise de sentimento, sumarização de textos

Processamento de Dados: OCR, expressões regulares, extração de dados de PDF

Ferramentas: Git, Jupyter Notebook, SAS

Idiomas

Português: Nativo

Inglês: Proficiência Profissional

Esperanto: Proficiência Profissional