

# Clébio Marques de Oliveira Júnior

clebiomojunior@gmail.com | medium.com/@juniorcl | linkedin/in/clebiojunior | github.com/juniorcl

## Resumo Profissional

Cientista de Dados com mais de **4 anos de experiência** em **machine learning, NLP e análise de dados** para solucionar problemas de negócio com impacto real.

Experiência em projetos de **score de crédito, previsão de vendas, segmentação de clientes e análise de comportamento**, além de extração e tratamento de **dados não estruturados** com técnicas avançadas de OCR e regex.

Atuação em empresas de tecnologia e consultoria (Vert Analytics, Datarisk, Be.X!), com entregas envolvendo desde **exploração e modelagem de dados** até a **construção de dashboards interativos em SAS**.

Mantenho portfólio ativo no GitHub e blog técnico no Medium, além de contribuições reconhecidas no Kaggle.

## Experiência Profissional

### Cientista de Dados – *Vert Analytics*

*Out 2024 – Presente*

- Desenvolvimento de dashboards estratégicos em **SAS**, a partir da exploração de dados estruturados e não estruturados.
- **Extração avançada de dados não estruturados** (PDF, OCR, regex), tratamento de missing values e integração em análises temporais.
- Projeto de **NLP para análise de comentários em redes sociais**, usando **spaCy** (tokenização, lematização, stopwords, correção gramatical), com resultados exibidos em dashboards interativos.

### Cientista de Dados – *Datarisk*

*Jan 2022 – Ago 2024*

- **Score de crédito**: modelo de ML que aumentou a precisão na concessão de crédito, reduzindo risco de inadimplência.
- **Previsão de vendas**: modelo de regressão que reduziu o erro médio em 12 %, apoiando planejamento estratégico.
- **Clusterização de clientes**: segmentação de perfis, viabilizando campanhas de marketing mais direcionadas.
- **Modelos preditivos de comportamento**: identificação de clientes com propensão ao churn, risco de inadimplência e instabilidade no emprego.

### Cientista de Dados – *Be.X!*

*Mar 2021 – Jan 2022*

- Extração de informações de **dados não estruturados** com regex e NLP.

- Detecção de **outliers com regras de negócio**, reduzindo impactos em relatórios operacionais.
- Predição de **atrasos em entregas** via modelos de ranking, permitindo ações preventivas na logística.

## Formação Acadêmica

- **Mestrado em Ciências Naturais (Ciências Ambientais)** – 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: extraindo padrões de dados
- Reconhecimento de Textos com OCR e Python
- Machine Learning e Data Science com Python de A a Z

## Habilidades Técnicas

- **Linguagens & Ferramentas:** Python (pandas, scikit-learn, spaCy, pdfplumber, pytesseract), SQL, Git, SAS, Jupyter
- **Machine Learning:** regressão, classificação, clustering, redes neurais, análise temporal
- **NLP:** tokenização, lematização, regex, sumarização, análise de sentimento
- **Outros:** OCR, dashboards interativos, deploy de modelos, visualização de dados

## Idiomas

- Português – Nativo
- Inglês – Profissional
- Esperanto – Profissional