

## **SÉRGIO LUIZ DOS SANTOS**

### **Cientista de Dados | Produtos Analíticos Orientados a Risco | Machine Learning em Produção**

Nova Lima, MG — Brasil

Email: [santossergiorealbjj@outlook.com](mailto:santossergiorealbjj@outlook.com)

Telefone: +55 31 99370-8399

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

Portfólio: <https://portfoliosantossergio.vercel.app>

### **RESUMO PROFISSIONAL**

Cientista de Dados focado na construção de produtos analíticos voltados à redução de risco e geração de impacto financeiro mensurável. Combina mais de 20 anos de experiência em sistemas regulados e de missão crítica com desenvolvimento completo de soluções de machine learning.

Experiência em traduzir problemas de negócio em soluções orientadas por dados ao longo de todo o ciclo de vida de ML: definição do problema, análise exploratória de dados (EDA), engenharia de atributos, treinamento de modelos, validação, implantação e monitoramento contínuo.

Forte capacidade de conectar desempenho estatístico à tomada de decisão de negócio, especialmente em ambientes sensíveis a risco e conformidade regulatória.

Inglês: Intermediário-Avançado (leitura técnica avançada; aprimorando fluência conversacional)

### **HABILIDADES TÉCNICAS**

#### **Linguagens de Programação**

Python | SQL | C# | .NET

## **Ciência de Dados & Machine Learning**

Pandas | NumPy | SciPy | Scikit-learn

Gradient Boosting (XGBoost, LightGBM – projetos exploratórios)

Aprendizado Supervisionado (Regressão, modelos baseados em árvores)

Análise Exploratória de Dados (EDA)

Engenharia de Atributos

Técnicas de validação cruzada

Métricas de avaliação de modelos (MAE, RMSE,  $R^2$ )

Seleção de modelos e análise de trade-offs

Compreensão do viés-variância

Prevenção de overfitting

Prevenção de vazamento de dados (data leakage)

Conhecimento básico de Deep Learning (TensorFlow, PyTorch – estudos acadêmicos)

Fundamentos de inferência causal e limitações de modelos preditivos

## **Dados & Infraestrutura em Nuvem**

Azure Databricks | Supabase | Render | Docker

Windows Server | Linux Server

## **Visualização de Dados**

Power BI | Matplotlib | Seaborn

## **Governança & Ambientes Críticos**

Rastreabilidade de dados

Controle de acesso

Continuidade operacional

Documentação técnica

Atuação em ambientes regulados

## **PROJETO EM DESTAQUE**

### **Plataforma Preditiva de Risco Operacional**

GitHub: [github.com/Santosdevbjj/analiseRiscosAtrasoObras](https://github.com/Santosdevbjj/analiseRiscosAtrasoObras)

Projeto desenvolvido com dados sintéticos estruturados, baseados em premissas operacionais realistas.

#### **Problema de Negócio**

Aproximadamente 35% das entregas eram concluídas com atraso, gerando penalidades contratuais e custos operacionais indiretos. As decisões eram reativas, tomadas apenas após a ocorrência do atraso.

Pergunta central:

Quais operações apresentam maior risco e onde a empresa deve agir primeiro para reduzir a exposição financeira?

#### **Linha de Base (Baseline)**

Previsões baseadas apenas na média histórica de atraso.

Erro Médio Absoluto (MAE): 12 dias.

Ausência de mecanismo estruturado de priorização de risco.

#### **Estratégia de Solução**

- Estruturação de arquitetura de dados em camadas (raw → analytics → products)

- Consolidação de variáveis operacionais (clima, tipo de solo, avaliação de fornecedor, cronograma)
- Engenharia de atributos orientada a indicadores de risco
- Tratamento de valores ausentes e padronização de variáveis categóricas
- Aplicação de validação cruzada para robustez
- Comparação entre Regressão Linear e RandomForestRegressor
- Avaliação de trade-offs entre interpretabilidade e performance preditiva

O RandomForest foi selecionado devido à sua robustez frente a variáveis heterogêneas e resiliência a dados ruidosos.

Riscos avaliados durante a modelagem:

- Mitigação de overfitting por meio da estratégia de validação
- Verificações de vazamento de dados na construção de atributos
- Consciência sobre relações não causais em modelos preditivos

### **Performance Técnica**

MAE (Baseline): 12 dias

MAE (Modelo): 4,97 dias

Redução de erro: 59%

Redução consistente da incerteza em ambiente operacional de alta variabilidade.

### **Performance de Negócio**

Conversão do erro estatístico em exposição financeira:

Redução estimada de incerteza: ~60%

Impacto financeiro anual estimado:  $\approx$  R\$ 248.400

Mudança de decisão: Reativa  $\rightarrow$  Preventiva

Cada erro de previsão foi convertido em risco financeiro residual, permitindo priorização estruturada e intervenção proativa.

### **Implantação em Produção**

- Modelo implantado em ambiente de nuvem (Render)
- API REST implementada com FastAPI
- Modelo containerizado com Docker
- Versionamento de modelo e práticas de reprodutibilidade durante a implantação
- Bot no Telegram para consultas operacionais com geração automática de relatórios em PDF
- Simulador executivo desenvolvido em Streamlit
- Camada de acesso a dados gerenciada via Supabase

Foco em acessibilidade para stakeholders não técnicos, rastreabilidade e usabilidade em ambiente real.

### **Principal Aprendizado**

Converter métricas de erro de modelo em impacto financeiro defensável e comunicar claramente trade-offs estatísticos para tomadores de decisão em ambientes sensíveis a risco.

## **EXPERIÊNCIA PROFISSIONAL**

### **Cientista de Dados – Projetos Aplicados**

Autônomo | 2023 – Presente

- Execução completa do ciclo de vida de ML: definição do problema → EDA → engenharia de atributos → validação → implantação → monitoramento
- Desenvolvimento de produtos analíticos prontos para produção
- Versionamento de modelos e práticas de reprodutibilidade em ambientes produtivos
- Tradução de desempenho estatístico em impacto mensurável de negócio

- Documentação alinhada a práticas de governança e rastreabilidade de dados
- Avaliação de trade-offs, limitações e riscos operacionais dos modelos

### **Consultor em Sistemas Críticos & Infraestrutura**

Autônomo | 2008 – Presente

- Consultoria em TI em ambientes corporativos de missão crítica
- Administração de servidores Windows e Linux
- Integração de sistemas legados com soluções modernas
- Suporte a departamentos jurídicos e administrativos
- Atuação em contextos regulados e sensíveis à conformidade

### **Banco Bradesco S.A.**

Escriturário → Programador → Analista de Sistemas | 1985 – 2008

Evolução de carreira desde operações administrativas até análise de sistemas e liderança técnica em instituição bancária de grande porte.

Principais Resultados:

- Automação de processos internos, eliminando aproximadamente 2.920 horas/ano de trabalho manual
- Suporte a sistemas utilizados por mais de 500 usuários com 99,5% de disponibilidade
- Desenvolvimento de sistemas com foco em rastreabilidade, confiabilidade e conformidade regulatória
- Atuação em ambiente bancário altamente regulado e exposto a risco operacional

### **FORMAÇÃO & CERTIFICAÇÕES**

Data Science com Python – 2025

Azure Databricks – Microsoft – 2025

Neo4J e Análise de Grafos – 2025

IBM AI Fundamentals – 2024

Microsoft Azure Bootcamp – 2024

Power BI e Excel Avançado – 2024

### **STACK ATUAL**

Python | SQL | Pandas | NumPy | Scikit-learn | XGBoost | Azure Databricks | Power BI | Neo4J | Docker | Streamlit | Supabase