

SERGIO LUIZ DOS SANTOS

**Científico de Datos | Productos Analíticos
Orientados al Riesgo | Machine Learning en
Producción**

Nova Lima, MG — Brasil

Email: santossergiorealbjj@outlook.com

Teléfono: +55 31 99370-8399

LinkedIn:

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

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

RESUMEN PROFESIONAL

Científico de Datos enfocado en la construcción de productos analíticos orientados a la reducción de riesgo y al impacto financiero medible. Combina más de 20 años de experiencia en sistemas regulados y de misión crítica con el desarrollo integral de soluciones de machine learning.

Experiencia en la traducción de problemas de negocio en soluciones basadas en datos a lo largo de todo el

ciclo de vida de ML: definición del problema, análisis exploratorio de datos (EDA), ingeniería de características, entrenamiento de modelos, validación, despliegue y monitoreo continuo.

Sólida capacidad para conectar el desempeño estadístico con la toma de decisiones empresariales, especialmente en entornos sensibles al riesgo y al cumplimiento normativo.

Inglés: Intermedio-Avanzado (lectura técnica avanzada; mejorando fluidez oral)

HABILIDADES TÉCNICAS

Lenguajes de Programación

Python | SQL | C# | .NET

Ciencia de Datos y Machine Learning

Pandas | NumPy | SciPy | Scikit-learn

Gradient Boosting (XGBoost, LightGBM – proyectos exploratorios)

Aprendizaje Supervisado (Regresión, modelos basados en árboles)

Análisis Exploratorio de Datos (EDA)

Ingeniería de Características

Técnicas de validación cruzada

Métricas de evaluación de modelos (MAE, RMSE, R^2)

Selección de modelos y análisis de trade-offs

Comprensión del equilibrio sesgo-varianza

Prevención de overfitting

Prevención de fuga de datos (data leakage)

Conocimientos básicos de Deep Learning (TensorFlow, PyTorch – estudios académicos)

Fundamentos de inferencia causal y limitaciones del modelado predictivo

Datos e Infraestructura en la Nube

Azure Databricks | Supabase | Render | Docker

Windows Server | Linux Server

Visualización de Datos

Power BI | Matplotlib | Seaborn

Gobernanza y Entornos Críticos

Trazabilidad de datos

Control de acceso

Continuidad operativa

Documentación técnica

Trabajo en entornos regulados

PROYECTO DESTACADO

Plataforma Predictiva de Riesgo Operacional

GitHub:

github.com/Santosdevbjj/analiseRiscosAtrasoObras

Proyecto desarrollado utilizando datos sintéticos estructurados basados en supuestos operacionales realistas.

Problema de Negocio

Aproximadamente el 35% de las entregas se completaban con retraso, generando penalizaciones contractuales y costos operativos indirectos. Las decisiones eran reactivas, tomadas únicamente después de que ocurrían los retrasos.

Pregunta central:

¿Qué operaciones presentan mayor riesgo y dónde debería actuar primero la empresa para reducir la exposición financiera?

Línea Base (Baseline)

Pronósticos basados únicamente en el promedio histórico de retraso.

Error Absoluto Medio (MAE): 12 días.

Ausencia de un mecanismo estructurado de priorización de riesgo.

Estrategia de Solución

- Diseño de arquitectura de datos en capas (raw → analytics → products)

- Consolidación de variables operativas (clima, tipo de suelo, calificación de proveedores, cronograma)
- Ingeniería de características orientada a indicadores de riesgo
- Tratamiento de valores faltantes y estandarización de variables categóricas
- Aplicación de validación cruzada para garantizar robustez
- Comparación entre Regresión Lineal y RandomForestRegressor
- Evaluación de trade-offs entre interpretabilidad y desempeño predictivo

RandomForest fue seleccionado debido a su robustez frente a variables heterogéneas y su resiliencia ante datos ruidosos.

Riesgos evaluados durante el modelado:

- Mitigación de overfitting mediante estrategia de validación
- Verificación de fuga de datos durante la construcción de características
- Conciencia sobre relaciones no causales en modelos predictivos

Desempeño Técnico

MAE (Baseline): 12 días

MAE (Modelo): 4,97 días

Reducción del error: 59%

Reducción consistente de la incertidumbre en un entorno operacional de alta variabilidad.

Desempeño de Negocio

Conversión del error estadístico en exposición financiera:

Reducción estimada de incertidumbre: ~60%

Impacto financiero anual estimado: ≈ R\$ 248.400

Cambio en la toma de decisiones: Reactiva → Preventiva

Cada error de predicción fue convertido en riesgo financiero residual, permitiendo priorización estructurada e intervención proactiva.

Despliegue en Producción

- Modelo desplegado en entorno cloud (Render)
- API REST implementada utilizando FastAPI
- Modelo containerizado con Docker
- Prácticas de versionado de modelos y reproducibilidad durante el despliegue
- Bot de Telegram para consultas operativas con generación automática de reportes en PDF
- Simulador ejecutivo desarrollado en Streamlit
- Capa de acceso a datos gestionada mediante Supabase

Enfoque en accesibilidad para stakeholders no técnicos, trazabilidad y usabilidad en entorno real.

Aprendizaje Clave

Traducir métricas de error de modelos en impacto financiero defendible y comunicar claramente los trade-offs estadísticos a tomadores de decisión en entornos sensibles al riesgo.

EXPERIENCIA PROFESIONAL

Científico de Datos – Proyectos Aplicados

Independiente | 2023 – Presente

- Ejecución integral del ciclo de vida de ML: definición del problema → EDA → ingeniería de características → validación → despliegue → monitoreo
- Desarrollo de productos analíticos listos para producción
- Prácticas de versionado de modelos y reproducibilidad en entornos productivos
- Traducción del desempeño estadístico en impacto de negocio medible
- Documentación alineada con prácticas de gobernanza y trazabilidad de datos
- Evaluación de trade-offs, limitaciones y riesgos operativos de los modelos

Consultor en Sistemas Críticos e Infraestructura

Independiente | 2008 – Presente

- Consultoría en TI en entornos corporativos de misión crítica
- Administración de servidores Windows y Linux
- Integración de sistemas legados con soluciones modernas
- Soporte a departamentos legales y administrativos
- Trabajo en contextos regulados y sensibles al cumplimiento normativo

Banco Bradesco S.A.

Empleado Administrativo → Programador → Analista de Sistemas | 1985 – 2008

Progresión de carrera desde operaciones administrativas hasta análisis de sistemas y liderazgo técnico en institución bancaria de gran escala.

Resultados Clave:

- Automatización de procesos internos, eliminando aproximadamente 2.920 horas/año de trabajo manual

- Soporte a sistemas utilizados por más de 500 usuarios con 99,5% de disponibilidad
- Desarrollo de sistemas enfocados en trazabilidad, confiabilidad y cumplimiento regulatorio
- Trabajo en entorno bancario altamente regulado y con exposición a riesgo operacional

EDUCACIÓN Y CERTIFICACIONES

Data Science con Python – 2025

Azure Databricks – Microsoft – 2025

Neo4J y Análisis de Grafos – 2025

IBM AI Fundamentals – 2024

Microsoft Azure Bootcamp – 2024

Power BI y Excel Avanzado – 2024

STACK ACTUAL

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